

TESIS DOCTORAL

**Arquitectura de la cromatina de los complejos *Six* en el
desarrollo y la evolución**

Autor: Carlos Gómez Marín

Director: Prof. Dr. Jose Luis Gómez Skarmeta

Tutor: Antonio J. Pérez Pulido

Universidad Pablo de Olavide

Centro Andaluz de Biología del Desarrollo

Consejo Superior de Investigaciones Científicas



A Rocio y Carlos

A mi familia

Q: What are the most useful attributes in today's scientific world?

A: That's easy: inexhaustible optimism, inexhaustible curiosity, inexhaustible energy and inexhaustible honesty!

Eric H. Davidson

ABSTRACT

The vast morphological variability observed across the animal kingdom is thought to be mainly achieved by modifying the expression patterns of a common toolbox of genes during development. Cis regulatory elements (CREs) are small DNA sequences scattered throughout the non-coding genome and are responsible for the spatio-temporal control of transcription through physical interaction with core promoters. They harbor binding sites for proteins such as transcription factors and chromatin organizers like CTCF and cohesin. The chromatin of multiple bilaterian species is compartmentalized in topologically associating domains (TADs), genomic regions within which sequences preferentially contact each other. This organization has been proposed to be essential to organize the regulatory information contained in animal genomes, bringing together CREs and their target promoters and avoiding promiscuous contacts. However, the impact of this spatial chromatin organization on gene expression, its degree of evolutionary conservation and the molecular mechanisms responsible for its formation are still poorly understood. In this thesis we have tried to contribute to the understanding of these aspects of chromatin spatial organization through the analysis of chromatin architecture of a specific locus: the Six cluster. Six homeobox genes are essential developmental regulators organized in gene clusters conserved during evolution. By using 4C-seq technologies on the Six genes presents in zebrafish, mouse and sea urchin we reveal that the Six clusters share a deeply evolutionarily conserved 3D chromatin organization that predates the Cambrian explosion. This chromatin architecture generates two largely independent regulatory landscapes (RLs) contained in two adjacent TADs and it is maintained stable during development. By disrupting the conserved TAD border in one of the zebrafish Six clusters, we demonstrate that this border is critical for preventing competition between promoters and enhancers located in separated RLs, thereby generating different expression patterns in genes located in close genomic proximity. Moreover, evolutionary comparison of Six-associated TAD borders reveals the presence of CCCTC-binding factor (CTCF) sites with diverging orientations in all studied deuterostomes. Genome-wide examination of mammalian HiC data reveals that this conserved CTCF configuration is a general signature of TAD borders, underscoring that common organizational principles underlie TAD compartmentalization in deuterostome evolution.

Indice de Contenido

Introducción	1
1. EL GENOMA REGULADOR	3
2. ELEMENTOS CIS-REGULADORES: IDENTIFICACIÓN Y ASOCIACIÓN	8
2.1 EL PROCESO DE ACTIVACIÓN TRANSCRIPCIONAL EN METAZOOS	8
2.2 IDENTIFICACIÓN A ESCALA GENÓMICA DE ELEMENTOS REGULADORES	10
2.3 ASOCIACIÓN DE ELEMENTOS REGULADORES Y PROMOTORES DIANA	15
3. ORGANIZACIÓN TOPOLÓGICA DEL GENOMA Y SU IMPLICACIÓN EN LA REGULACIÓN GÉNICA	25
3.2 DOMINIOS TOPOLÓGICAMENTE ASOCIADOS (TADs)	26
3.3 PROTEÍNAS ARQUITECTÓNICAS, INTERACCIONES CONSTITUTIVAS Y BORDES TOPOLÓGICOS.	27
4. LOS GENES <i>Six</i>	31
Objetivos	37
Materiales y Métodos	41
1. DETERMINACIÓN DEL ENTORNO REGULADOR DE UN GEN: 4C-SEQ.	43
1.1 Preparación de muestras	43
1.2 Diseño de cebadores	48
1.3 PCRs y preparación de librerías de 4C	49
1.4 Procesamiento de los datos	51
1.5 Cálculo del borde topológico predicho (btp).	53
2. SELECCIÓN DE POTENCIALES ELEMENTOS REGULADORES EN EL COMPLEJO <i>SIX2A-SIX3A</i>	54
3. ENSAYO DE ELEMENTOS REGULADORES EN PEZ CEBRA	54
4. MODIFICACIÓN Y TRANSGÉNESIS DE BACS	57
4.1 Generación del BAC DK74B2-six2a::GFP-six3a::mCherry-iToI2. (FullBAC)	59
4.2 Generación del BAC Δ 18kb DK74B2-six2a::GFP-six3a::mCherry-iToI2 (DelBAC)	61
4.3 Purificación y transgénesis	62
5. BÚSQUEDA DE SITIOS DE UNIÓN DE CTCF	62
6. HIBRIDACIÓN IN SITU	64
6.1 Pez cebra	64
6.2 Ratón	65

6.3 Erizo de mar	65
Resultados	69
1. CARACTERIZACIÓN DEL ENTORNO REGULADOR DE LOS GENES <i>six2a</i> Y <i>six3a</i> EN EMBRIONES DE PEZ ZEBRA A 24 HPF.	71
2. ENTORNO REGULADOR DE LOS GENES <i>six3a</i> Y <i>six2a</i> DURANTE EL DESARROLLO DEL PEZ ZEBRA.	77
3. ORGANIZACIÓN TRIDIMENSIONAL DEL RESTO DE COMPLEJOS DE GENES <i>Six</i> DE PEZ ZEBRA	83
4. ORGANIZACIÓN TRIDIMENSIONAL DE LOS COMPLEJOS <i>Six</i> EN RATÓN	90
5. ORGANIZACIÓN TRIDIMENSIONAL DEL COMPLEJO <i>Six</i> EN ERIZO DE MAR	95
6. CONTRIBUCIÓN FUNCIONAL DEL BORDE TOPOLÓGICO PRESENTE EN EL COMPLEJO <i>six3a-six2a</i> DE PEZ CEBRA.	97
7. BÚSQUEDA DE LOS DETERMINANTES DE LA ORGANIZACIÓN TRIDIMENSIONAL DE LOS COMPLEJOS <i>Six</i>	102
Discusión	114
1. DE LA GENÓMICA COMPARATIVA A LA ARQUITECTÓMICA COMPARATIVA.....	116
2. ORGANIZACIÓN 3D DE LOS COMPLEJOS SIX: PEZ CEBRA.....	120
3. ORGANIZACIÓN 3D DE LOS COMPLEJOS SIX: UNA MIRADA EVOLUTIVA	123
4. LA FUNCION DEL BORDE TOPOLÓGICO	126
5. UNA POSIBLE HUELLA MOLECULAR EN LOS BORDES TOPOLÓGICOS	129
6. INTERACCIONES ESTRUCTURALES, ORGANIZACIÓN 3D DE LA CROMATINA, PLEIOTROPIA Y EVOLUCIÓN.	133
Conclusiones	140
Anexos	144
1. ANEXO 1: Protocolo 4C-seq	145
2. ANEXO 2: Protocolo ChIP-4C.	149
3. ANEXO 3: Cebadores 4C-seq.....	151
4. ANEXO 4: Regiones usadas para testar actividad potenciadora	152
5. ANEXO 5: Cebadores para la preparación de sondas de los genes <i>Six</i> de pez cebra	152
6. ANEXO 6: Líneas transgénicas estables de las versiones del BAC DKEY-74B2, FullBAC (24hpf, 48hpf, 5dpf) y Δ 18kbBAC (5 dpf).	153
A) 24 hpf	153
B) 48 hpf	154
C) 5 dpf	155
Bibliografía	158
Agradecimientos	175

Índice de Figuras

Introducción

1 FIG I_1	6
2 FIG I_2	13
3 FIG I_3	22
4 FIG I_4	29
5 FIG I_5	33

Materiales y Métodos

6 FIG. MYM_1	48
7 FIG MYM_2	50
8 FIG MYM_3	52
9 FIG MYM_4	57
10 FIG MYM_5	59

Resultados

11 FIG RE_1	72
12 FIG RE_2	75
13 FIG RE_	80
14 FIG RE_4	82
15 FIG RE_5	84
16 FIG RE_6	86
17 FIG RE_7	88
18 FIG RE_8	89
19 FIG RE_9	92
20 FIG RE_20	94
21 FIG RE_11	98
22 FIG RE_12	101
23 FIG RE_13	104
24 FIG RE_14	106
25 FIG RE 15	108
26 FIG RE_16	109

Discusión

27 FIG D_1	127
28 FIG D_2	135

Introducción

1. EL GENOMA REGULADOR

El plan corporal de los animales, sus estructuras y las funciones con las que sus morfologías los dotan, son el resultado de la suma, en el tiempo y el espacio, de los sucesivos procesos de desarrollo que tienen lugar en cada animal. En general, el desarrollo embrionario es mediado por la regulación espacial y temporal de miles de genes y consiste en una progresión dinámica de **“estados regulatorios”**. Un “estado regulatorio” se define por la presencia y estado de actividad en el núcleo celular de un conjunto de **factores de transcripción** (proteínas reguladoras capaces de reconocer y unirse a una secuencia de DNA), que determinan la expresión génica. Físicamente, la estructura encargada de interpretar estos “estados regulatorios” es la suma de secuencias modulares de ADN dispersas en el genoma, capaces de interactuar con los factores de transcripción. Estos módulos “leen” la información suministrada por el “estado regulatorio” de la célula y la “procesan” transformándola en instrucciones que pueden ser utilizadas por la maquinaria bioquímica para la expresión de genes específicos. Estos elementos de ADN capaces de “leer” y “procesar” el “estado regulatorio” de una célula son lo que llamamos elementos cis reguladores. El contenido en secuencia, la disposición y otros aspectos organizativos de estas **estructuras de control modular** son el patrimonio característico de cada especie. El conjunto de estos módulos determinará el resultado particular del proceso de desarrollo de cada embrión, y por tanto, su morfología y las funciones asociadas a ella. En la evolución, la alteración de los planes corporales se ha producido en gran medida a consecuencia de cambios en este **código genómico** para la regulación de genes durante el desarrollo (Davidson, 2006). Por otro lado, desde el punto de vista biomédico, los estudios genómicos de asociación de enfermedades humanas (del inglés Genome Wide Association Studies - GWAS-) muestran como la mayoría de los polimorfismos de un solo nucleótido (SNP) no están asociados a ninguna secuencia codificante. De hecho muchos se asocian con posibles elementos reguladores distales (Maurano y cols 2012). Estos aspectos ponen de manifiesto la importancia, no sólo científica, sino también económica y social que tiene la comprensión del “genoma regulador”.

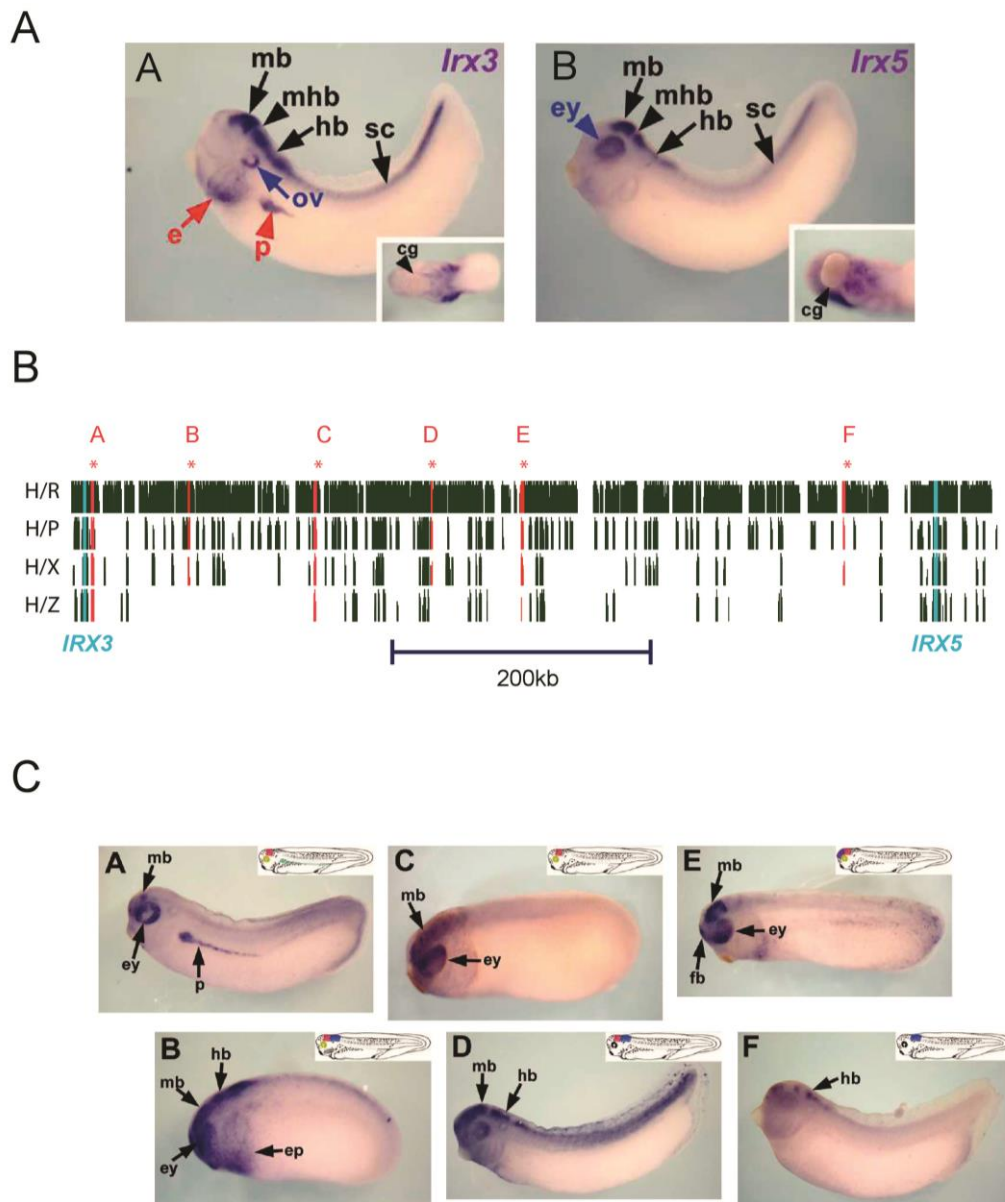
El estudio del genoma como entidad molecular que alberga el lenguaje de la vida ha sufrido un gran avance desde el desarrollo de las nuevas tecnologías de secuenciación masiva de ADN. La aplicación de esta tecnología a los genomas de multitud de animales ha revelado

que la complejidad y diversidad morfológica y celular de un organismo no puede asociarse exclusivamente al número de genes. Los mamíferos poseen una complejidad morfológica mayor que los nematodos, sin embargo, ambos contienen alrededor de 20000 genes. Además si la comparación de los genes se centran en aquellos que codifican para factores de transcripción y componentes de las rutas de señalización necesarias para controlar y dirigir el proceso de desarrollo no existen apenas diferencias cualitativas entre organismos bilaterales (Erwin. 2009, Putnam y cols 2007). Deben existir, por tanto, otras formas de generar la diversidad morfológica que encontramos en el reino animal. Algunos parámetros que contribuyen al aumento de las diferencias entre especies son las posibles modificaciones postraduccionales de las proteínas o la variabilidad en el *splicing* alternativo de los genes (Keren y cols 2010). Por otro lado, la gran diversidad morfológica del mundo animal se ha producido en base a un reducido y conservado número de genes de desarrollo. Esto ha sido posible porque muchos de estos genes han adquirido múltiples funciones durante el desarrollo embrionario, interviniendo en distintos momentos y estructuras de dicho proceso. Este fenómeno se conoce como pleiotropía. Algunas funciones de este grupo de genes son comunes a muchas especies, otras aparecen de forma específica en una o varias especies en particular. La posibilidad de que un gen pueda ser usado en un territorio o momento específico del desarrollo, dependerá de la capacidad del mismo de ser expresado en dicho momento y territorio. Si un gen de desarrollo adquiere la capacidad de expresarse en un nuevo tejido, un momento distinto o con una intensidad diferente podría modificar el resultado final del proceso de embriogénesis del organismo sin necesidad de cambiar el número de genes. El conjunto de territorios y momentos del desarrollo embrionario en los que un gen se expresa se conoce como “patrón de expresión” y viene determinado por la información reguladora a la que dicho gen está asociado. Físicamente, esta información reguladora la componen principalmente los elementos cis-reguladores. La modificación, adquisición o pérdida de estos elementos de ADN se ha postulado como una de las principales fuerzas evolutivas generadoras de diversidad morfológica.

El aumento por el interés en el estudio de estos elementos cis-reguladores se produjo cuando, a partir de la secuenciación y comparación de distintos genomas animales y en particular de vertebrados, se comprobó que la proporción del genoma dedicada a la codificación de proteínas estaba en torno al 3-5%, encontrando incluso desiertos génicos (regiones desprovistas de secuencias codificantes de proteínas) que llegan a expandirse varias megabases (Venter y cols 2001, Nobrega y cols 2003, Montavon y Duboule 2012). La

genómica comparativa entre distintas especies de vertebrados reveló la existencia de multitud de pequeños módulos de ADN dispersos en las regiones no codificantes que estaban conservados entre distintas especies, siendo su número mayor cuanto mayor era la cercanía filogenética de las especies comparadas (Tagle y cols 1988, Alonso y cols 2009). A estos módulos se los denominó HCNRs, del inglés “Highly Conserved Non-coding Regions”. La gran mayoría de estos módulos se encontraban flanqueando genes de desarrollo que mostraban patrones de expresión complejos. Una gran proporción de estos HCNRs mostraban además actividad potenciadora de la transcripción mediante ensayos de transgénesis en ratón, *Xenopus* o pez cebra, llevando la expresión de genes reporteros a territorios consistentes con la expresión endógena del gen flanqueado (de la Calle-Mustienes y cols 2005, Wolfe y cols 2005, McEwen y cols 2006). Por ejemplo, en el trabajo de de la Calle-Mustienes y cols se exploraron los desiertos génicos que flanquean los complejos de los genes de desarrollo *Irx* (Fig. I_1, B). Mediante genómica comparativa entre distintas especies de vertebrados (humano, ratón, rana y pez) se detectaron multitud de HCNRs que al ser ensayados mediante transgénesis en *Xenopus* mostraron la capacidad de activar la transcripción de genes reporteros en territorios que solapaban con la expresión endógena de los genes *Irx* (Fig. I_1, C).

Podemos encontrar elementos cis-reguladores con actividad “potenciadora”, “silenciadora” y/o “aisladora”. Un elemento cis-regulador con actividad potenciadora se define como una porción de ADN capaz de activar la transcripción desde un promotor diana independientemente de la orientación o la distancia entre el potenciador y el promotor. Estos elementos de ADN contienen sitios de unión para factores de transcripción generales y específicos, por lo que se piensa que constituyen un foco de unión para dichos factores (Banerji y cols 1983; Gilies y cols 1983; Buecker y Wysocka 2012). Un elemento cis-regulador con actividad silenciadora se corresponde con el elemento opuesto, aquel capaz de inactivar la transcripción. La cooperación entre múltiples elementos cis-reguladores potenciadores y/o silenciadores puede ser de tipo aditiva, sinérgica o competitiva. Otro elemento regulador distal, serían los conocidos como “aisladores”, capaces de establecer dominios transcripcionales discretos. Estos elementos tienen la capacidad de bloquear la acción de los elementos cis-reguladores sobre los promotores, y/o la capacidad de actuar como barrera a la expansión de la heterocromatina. La mayoría de los aisladores de vertebrados contienen sitios de unión de la proteína CTCF (Yusufzai y Felsenfeld 2004; Cuddapah y cols 2009).



Adaptado de de la Calle-Muestienes y cols 2005

1 Fig I_1

Fig. I_1: Los desiertos génicos se encuentran, con frecuencia, poblados de elementos cis-reguladores: **A)** Expresión de los genes *Irx3* e *Irx5* en embriones de *Xenopus* (mb: cerebro medio; mhb: borde medio-posterior del cerebro; hb: cerebro posterior; sc: espina cordal; ey: ojo; ov: vesícula ótica; e: capa ectodérmica de las branquias; p: pronefros; cg: glándula de cemento). **B)** Genómica comparativa de la región IRX3-IRX5 de humano (H) con: ratón(R), pollo(P), rana(X) y pez cebra(Z). Las regiones verdes representan secuencias conservadas no codificantes, las azules regiones conservadas codificantes y las rojas (A,B,C,D,E,F) regiones conservadas no codificantes que mostraron actividad potenciadora en ensayos de transgénesis. **C)** Actividad potenciadora de los elementos A,B,C,D,E, y F en embriones de *Xenopus*.

Uno de los aspectos más críticos en el estudio de los elementos cis-reguladores es la gran complejidad que tiene su identificación. Hemos visto que los HCNRs hacen referencia a elementos cis-reguladores que están altamente conservados entre distintas especies. Sin embargo no todos los elementos cis-reguladores de un organismo tienen por qué estar conservados en más de una especie ni todos los elementos conservados son elementos cis-reguladores. Como hemos mencionado, los elementos cis-reguladores están formados por secuencias que son reconocidas por un conjunto de proteínas de unión al ADN, principalmente factores de transcripción. Estos factores de transcripción suelen reconocer secuencias de poco tamaño, que por azar se pueden encontrar frecuentemente en el genoma. Además los sitios de unión para un factor de transcripción concreto, pueden tener un alto grado de degeneración y un factor de transcripción puede unirse de manera promiscua a multitud de secuencias diferentes. Estos hechos hacen que exista poca correlación entre la existencia de una secuencia consenso de unión a un factor de transcripción particular y su capacidad real de reclutar a dicho factor. Por tanto la identificación a escala global de dichos elementos, en base exclusivamente a la secuencia genómica, se hace casi imposible. Finalmente, otra característica importante de estos elementos reguladores que dificulta su identificación es que pueden encontrarse a grandes distancias de su promotor diana, saltando en ocasiones a genes situados entre el elemento regulador y su diana. Además un mismo promotor puede estar controlado por multitud de elementos cis-reguladores, y un mismo elemento cis-regulador puede controlar la expresión de más de un gen. Dos aspectos que hacen que la asociación de elementos cis-reguladores con sus promotores diana se haya convertido en otro desafío de la genómica actual.

La gran importancia del genoma regulador (entendido como el conjunto total de elementos cis-reguladores presentes en el genoma de un organismo) en el control de la expresión génica en el desarrollo embrionario, en la historia evolutiva animal y en el desarrollo de enfermedades humanas (Erokhin y cols 2015, Wamstad y cols 2014, Dickel y cols 2013), junto con la complejidad para la identificación y asociación de los elementos reguladores que lo componen, han motivado que la comunidad científica haya realizado grandes esfuerzos durante las últimas décadas para el desarrollo de una gran variedad de nuevas tecnologías, procedimientos y técnicas para su estudio.

2. ELEMENTOS CIS-REGULADORES: IDENTIFICACIÓN Y ASOCIACIÓN

2.1 EL PROCESO DE ACTIVACIÓN TRANSCRIPCIONAL EN METAZOOS

Los primeros paradigmas de control génico se establecieron con el estudio de la transcripción en bacterias y fagos, que típicamente usan el promotor como único sitio de integración de la información necesaria para activar o reprimir un gen. Los primeros sistemas estudiados consistían en la unión de un agente represor a una secuencia solapante con el promotor que impedía la unión de la ARN polimerasa. Estos elementos represores podían separarse del ADN mediante la unión a un elemento o molécula inductora que inducía cambios conformacionales en el represor, permitiendo de este modo la unión de la ARN polimerasa a la secuencia promotora (revisado en Lewis 2013). Además de las secuencias promotoras se caracterizaron otros elementos importantes de la regulación génica procariota, como fue el descubrimiento de los elementos activadores y factores sigma, que mediante la unión a secuencias localizadas entre 50–60 pb del inicio de la transcripción, ayudaban a reclutar y estabilizar la ARN polimerasa en las secuencias promotoras (Losick 1998). En general, bacterias y fagos usan sólo elementos reguladores cercanos o incluidos en la región promotora para el control de la transcripción de sus genes.

Al final de la década de los 70, los científicos comenzaron a tener las primeras ideas sobre la organización de los genomas de los eucariotas. En comparación con las bacterias, se encontraron tres diferencias fundamentales: la presencia de intrones, que interrumpían las secuencias codificantes de los genes, la asociación del ADN a las histonas (dificultando el acceso de factores en trans), y por último la identificación de secuencias reguladoras en cis alejadas de la secuencia promotora a la que regulaban. En particular los metazoos, comparados con otros eucariotas, tienen genomas particularmente poco compactos, con tamaños de intrones y distancias intergénicas muy grandes.

Probablemente para responder a las inmensas demandas transcripcionales, en el tiempo y en el espacio, que requiere el desarrollo embrionario y la formación de miles de tipos celulares distintos, los metazoos contienen una maquinaria proteica encargada de la regulación de la transcripción mucho mayor y más compleja que las bacterias (Levine y Tijan 2003). No sólo por la aparición de multitud de factores de transcripción específicos de tejido, sino también

por una mayor complejidad y diversidad de los componentes de la maquinaria basal de la transcripción, también conocida como Complejo de Pre-iniciación (CPI) de la transcripción. El prototipo de CPI eucariota consiste en más de 85 polipéptidos, incluyendo varios complejos multiproteicos como la ARN polimerasa II, TFIID, E, F, H y varios coactivadores y remodeladores de la cromatina. En los últimos años se ha descubierto que muchos componentes y coactivadores del CPI no son ni generales ni universales. Muchos de estos componentes son, de hecho, específicos de distintos tipos celulares o de distintos grupos de genes. Además existe también diversidad en la secuencia promotora básica (por ejemplo, la caja TATA frente a la DPE, *Downstream Promoter Element*), que contribuye a la especificidad de la respuesta transcripcional de un gen ante la acción de un mismo grupo de elementos reguladores (Roeder 1996; Cramer 2002; Goodrich y Tijan 2010; Juven-Gershon y Kadonaga 2010; Danino y cols 2015). Estos elementos también contribuyen a determinar las respuestas transcripcionales específicas de cada tipo celular, desde la secuencia promotora.

Los análisis a escala genómica del proceso de desarrollo embrionario en los metazoos, también han revelado que muchos de los elementos reguladores en cis y promotores están marcados con anterioridad a su uso en el proceso de activación transcripcional en sí mismo. Esto consigue una activación puntual y adecuada de los genes necesarios en cada momento, en respuesta a un estímulo determinado (Lagha y cols 2012). A nivel de los elementos cis-reguladores, esto se consigue gracias a lo que se conoce como “factores pioneros”, que son factores de transcripción capaces de unirse a los elementos cis-reguladores y marcarlos antes de su activación, probablemente mediante el reclutamiento de enzimas modificadoras de histonas como CBP/p300 (Visel y cols 2009) o remodeladoras de la cromatina como SWI/SNF (Ronan y cols 2013), dejando dichos potenciadores accesibles para los subsecuentes factores de transcripción. Algunos ejemplos de estos factores pioneros son el gen FOXA, capaz de marcar potenciadores específicos de hígado durante el proceso de desarrollo del hígado de ratón antes de que sean activos (Gualdi y cols 1996; Cirillo y cols 2002); los factores inductores de pluripotencia OCT4 y SOX2 (Soufi y cols 2012, Wapinski y cols 2013), o el factor de transcripción Zelda en *Drosophila*, que interviene marcando la mayoría de potenciadores necesarios para la inducción de la expresión génica del cigoto antes de que esta ocurra (Harrison y cols 2011, Liang y cols 2008, Nien y cols 2011). A nivel de los promotores, este marcaje anterior a la activación, se lleva a cabo mediante el reclutamiento de la ARN polimerasa II pausada (Chen y cols 2013), u otros miembros del

CPI como TBP y TFIIA (Guglielmi y cols 2013). Este premarcaje bloquearía el ensamblaje de nucleosomas inhibidores en la secuencia promotora. En consecuencia, los genes “pausados” mostrarían una activación inmediata tras la inducción.

2.2 IDENTIFICACIÓN A ESCALA GENÓMICA DE ELEMENTOS REGULADORES

El desacoplamiento de la regulación transcripcional de la región promotora y la distribución del control regulador a lo largo de multitud de elementos distales, proporciona un excelente nivel de precisión y robustez en el control de la expresión génica.

Se han encontrado elementos reguladores en los intrones de los genes, varios cientos de pares de bases aguas arriba y aguas abajo del promotor (pero también a distancias de millones de pares de bases) en intrones de genes alejados de los que regulan, no relacionados o incluso en zonas codificantes y en cromosomas distintos (Visel y cols 2009; Williams y cols 2010; Dong y cols 2010). Esta variabilidad en su localización potencial y la heterogeneidad de su secuencia, convierte la identificación y asociación de estos elementos reguladores distales en todo un desafío para la genómica estructural.

El desarrollo de las tecnologías de secuenciación masiva ha incrementado enormemente la capacidad para detectar elementos potenciadores (Hawkins RD y cols, 2010). La asociación de enzimas remodeladoras de la cromatina (p300, CHD7, Brg1) o de marcas epigenéticas específicas de las histonas con los elementos cis-reguladores, ha permitido el mapeo a escala genómica de millones de estos elementos (tanto en genomas humanos como otros organismos).

El ADN se encuentra enrollado alrededor de octámeros de proteínas de histonas. Estos octámeros forman los nucleosomas. Cada nucleosoma está formado por dos copias de las histonas H2A, H2B, H3 y H4. Las colas de los extremos amino terminal de las histonas están altamente conservadas y se disponen hacia el exterior del nucleosoma. Estas colas están sometidas a un gran número de modificaciones postraduccionales como la acetilación, la metilación, la fosforilación o la ubiquitinación. Algunas de estas modificaciones están asociadas con regiones funcionales específicas como pueden ser promotores, secuencias codificantes o elementos cis reguladores. Estas modificaciones son “leídas” por proteínas

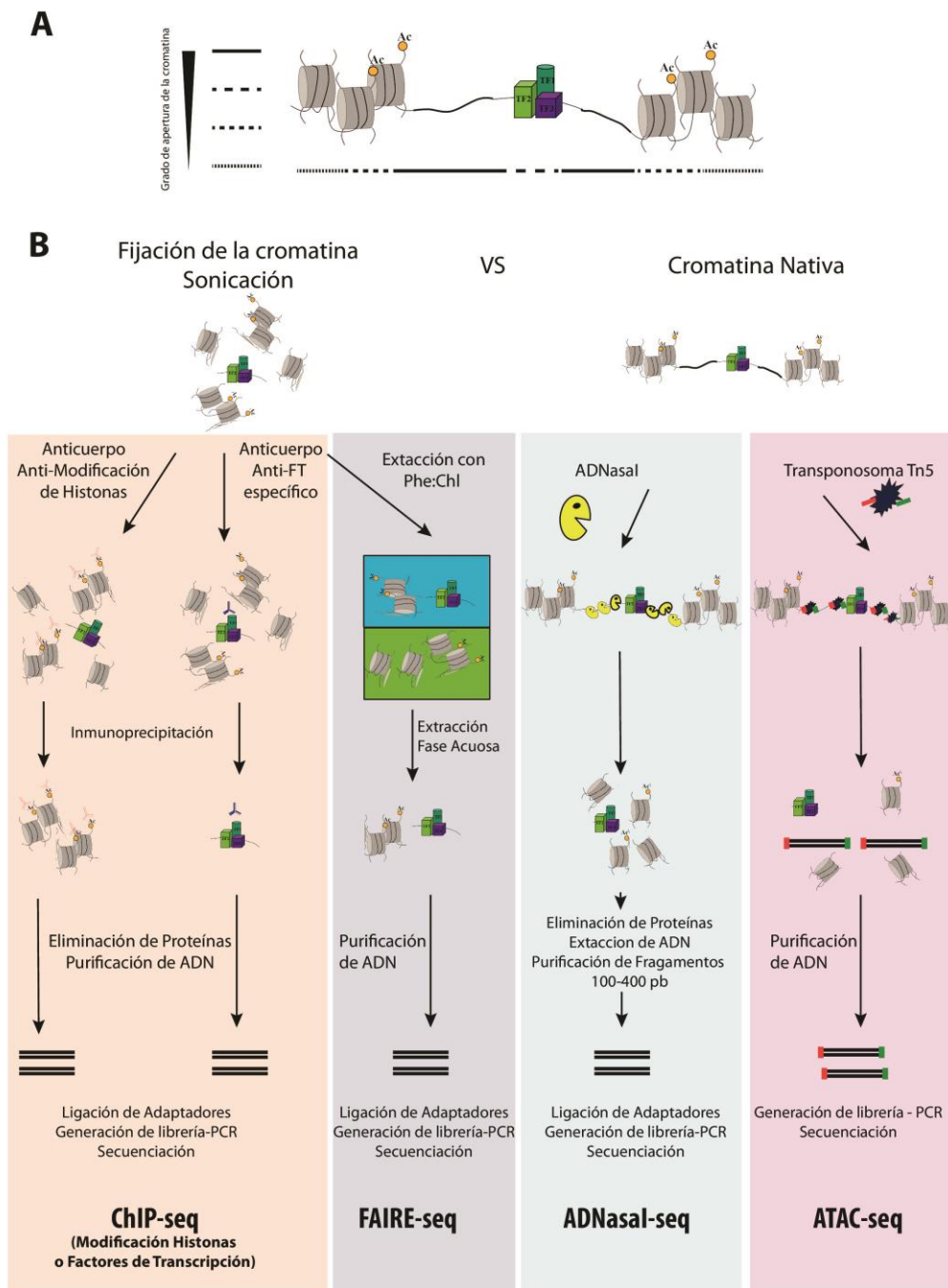
reguladoras, que contribuyen a determinar el grado de compactación de la cromatina o a reclutar otras proteínas que determinan el estado regulador (activo o inactivo) de esas regiones. Por tanto la detección a escala global de estas modificaciones ayuda a dibujar un mapa de los elementos funcionales del genoma en un contexto determinado. Además, dado que los elementos cis-reguladores “pausados” (pre-activos) o activos están asociados a nucleosomas altamente dinámicos, son susceptibles de ser detectados mediante técnicas que detecten las porciones del genoma libres de nucleosomas. A continuación pasamos a realizar una breve descripción de cada uno de estos métodos.

2.2.1 Métodos basados en la inmunoprecipitación de la cromatina

Las técnicas basadas en la inmunoprecipitación de la cromatina (ChIP), permiten obtener un perfil “*in vivo*” de las regiones de ADN unidas por proteínas específicas. En los experimentos de ChIP, todas las interacciones ADN-proteína son fijadas, las células lisadas y la cromatina fragmentada mediante sonicación en fragmentos de varios cientos de pares de bases. El uso de anticuerpos específicos contra proteínas o modificaciones postraduccionales de las mismas, permite la precipitación de un conjunto de fragmentos específicos de ADN. La identificación y cuantificación a escala genómica de esta librería de fragmentos de ADN precipitados se puede hacer mediante la hibridación de los mismos con microarrays de ADN (ChIP-chip) o mediante el uso de técnicas de secuenciación masiva (ChIP-seq).

La elección de la diana para el anticuerpo usado en este tipo de experimento determinará, por tanto, el tipo de información obtenida. El uso de anticuerpos contra factores de transcripción específicos, permitirá la identificación del subconjunto específico de elementos reguladores controlados por dichos factores de transcripción en el contexto experimental particular (tejido, tipo celular, tratamiento específico, etc.). Por otro lado si la diana para el anticuerpo es una proteína remodeladora de la cromatina o modificadora de histonas (p300, ATAC, CHD7, Brg1) asociadas al control de los elementos reguladores en general, se obtendría un mapa general de los elementos reguladores. Del mismo modo, se podría obtener este tipo de mapas generales mediante el uso de anticuerpos contra modificaciones postraduccionales de las histonas asociadas con el estado regulatorio específico de elementos cis-reguladores y promotores. Así, por ejemplo, la monometilación de la lisina 4 de la histona 3 (H3K4me1) está asociada a elementos reguladores distales y promotores en estado “pre-activado” (Guenther y cols 2007; Heintzman y cols 2007), mientras que la acetilación de la lisina 27 de

la histona 3 (H3K27ac) se ha asociado a elementos reguladores distales activos (Rada-Iglesias y cols 2011, Bogdanovic y cols 2012), y la trimetilación de la lisina 27 de la histona 3 (H3K27me3) se ha asociado con elementos reguladores reprimidos (Akkers y cols 2009, Lim y cols 2011). Por tanto, el uso de anticuerpos específicos contra este tipo de modificaciones de histonas, permite la obtención del conjunto de elementos reguladores que se encuentran en un estado regulador particular en las condiciones experimentales concretas.



2 Fig I_2

Fig. I_2: Métodos experimentales para la detección a escala genómica de elementos cis-reguladores. A) Representación esquemática de un elemento cis-regulador activo. Se caracteriza por la presencia de histonas con determinadas modificaciones postraduccionales como la H3K27ac, la unión de factores de transcripción así como por un mayor grado de apertura de la cromatina. Estos aspectos son aprovechados para el desarrollo de técnicas experimentales para la detección de elementos cis-reguladores. B) Principales etapas de los distintos procedimientos experimentales para la detección de elementos cis-reguladores a escala genómica: ChIP-seq, FAIRE-seq, ADNasal-seq y ATAC-seq. abierta”: FAIRE, ensayo de ADNasal y ATAC-seq.

2.2.2 Métodos basados en la detección de cromatina abierta

La unión de factores de transcripción a elementos reguladores está condicionada a la accesibilidad de la secuencia de ADN. Se ha visto que las regiones reguladoras interaccionan con nucleosomas híper-dinámicos, que contienen unas variantes específicas de histonas, H3.3 y H2A.Z, que se depositan de manera independiente a la replicación. Esto hace que estas regiones de la cromatina pasen un mayor tiempo desprovistas de nucleosomas y sean más sensibles a nucleasas de DNA (p.e: ADNasa I) o más permeables a transposasas (por ejemplo a la transposasa Tn5). Estas tres características físicas (desvinculación de los nucleosomas, sensibilidad a ADNasas y accesibilidad a transposasas) también se han usado como principio para la detección a escala genómica de elementos reguladores. Estas técnicas no requieren la utilización de ningún anticuerpo, por lo que pueden ser utilizadas en organismos donde la disponibilidad de dichos anticuerpos es limitada o en muestras donde su utilización es más compleja. Así encontramos los siguientes métodos de detección global de elementos reguladores basados en la “accesibilidad” de la cromatina (Fig. I_2):

- **ADNasaI-seq:** Esta técnica se basa en el uso de la endonucleasa ADNasaI. Esta enzima es capaz de digerir los fragmentos de ADN que no están asociados a histonas independientemente de su secuencia. A estos fragmentos se les conoce como “sitios hipersensibles a la ADNasaI” (DHSs, del inglés “DNAaseI hypersensitive sites”). Tras la digestión con la enzima, la purificación de los fragmentos digeridos se puede hacer bien, mediante la biotinilización de los extremos generados y la posterior purificación (Crawford y cols 2004) o bien directamente mediante la purificación de la porción de los fragmentos más pequeños (dado que la frecuencia de digestión es mucho mayor en la cromatina abierta) (Sabo y cols 2004). Tras esta purificación la librería de fragmentos se secuenciaría mediante técnicas de secuenciación masiva.
- **FAIRE:** Del inglés “Formaldehyde Assisted Isolation of Regulatory elements”. Esta técnica, al igual que el ChIP, se basa en la fijación de las interacciones proteínas-ADN mediante el uso de agentes fijadores tipo formaldehído. Tras la fijación, viene un paso de sonicación. Los fragmentos de ADN que no estuviesen asociados a nucleosomas, pueden ser separados de los que sí lo están, mediante un paso de fenolización. Esta librería enriquecida en fragmentos no asociados a nucleosomas puede secuenciarse mediante técnicas de secuenciación masiva, obteniéndose un

perfil de regiones de cromatina abierta (Giresi y cols 2007). Se trata de un proceso muy simple, capaz de detectar elementos reguladores distales.

- **ATAC-seq:** Del inglés “Assay for Transposase Accessible Chromatine using sequencing”. Esta técnica usa una transposasa bacteriana llamada Tn5. Esta transposasa ha sido ensamblada in vitro con unos adaptadores (cebadores) compatibles con las plataformas de secuenciación masiva de ADN, formando un transpososoma. La técnica se basa en el uso de cromatina “nativa” (no fijada) que se trata con esta Tn5 modificada. La Tn5 producirá una mayor frecuencia de transposición de los cebadores en las regiones desprovistas de nucleosomas (“activas”) que en las regiones cubiertas por los mismos. De esta forma esta Tn5 es capaz de fragmentar y marcar al mismo tiempo con inserciones un genoma completo (proceso conocido como “tagmentación”). La librería de fragmentos marcados por los cebadores de la Tn5 es amplificada por PCR y secuenciada. Además el uso de Tn5 cargada con secuencias modificadas de cebadores permite combinar varias muestras en un mismo proceso de secuenciación. Otra de las principales ventajas de esta técnica es que es capaz de llevarse a cabo con un número muy limitado de células (entre 500 y 50000 células frente a los varios millones que se necesitan en las técnicas anteriores). Por último esta técnica produce una gran resolución en la señal, lo que permite incluso identificar los sitios de unión de los distintos factores de transcripción (Buenrostro JD y cols, 2013).

2.3 ASOCIACIÓN DE ELEMENTOS REGULADORES Y PROMOTORES DIANA

Dada la enorme cantidad de elementos potenciadores y su capacidad de actuar a largas distancias, es fundamental poder asociar dichos elementos con sus promotores diana y conocer cuál es el mecanismo por el que son capaces de influir en la actividad de promotores que frecuentemente se encuentran a grandes distancias.

Originalmente se pensó que los factores de transcripción que se unían a los elementos potenciadores influían en la transcripción mediante el reclutamiento directo de la ARN polimerasa II y el CPI completo (Maston y cols 2006). Sin embargo, esto requeriría una

cercanía lineal mucho mayor entre el promotor y los elementos potenciadores. Hoy en día está generalmente aceptado que los potenciadores y los promotores interactúan mediante la formación de bucles de ADN. El primer estudio que puso de manifiesto el establecimiento de bucles de ADN como forma de regulación de la activación génica, fue llevado a cabo en *E. coli*, e incluía la represión del gen *AraC* y la activación a distancia del gen *NtrC* (Ptashne 1986). Otro de los estudios pioneros que mostraron como un bucle entre un elemento regulador y un promotor era necesario para la activación de este último, fue la activación del gen PRL (prolactina) por el receptor de estrógenos (Cullen 1993). En este caso el estrógeno, a través de su receptor, estimulaba la interacción entre el elemento regulador distal (200 pb) y el promotor, 3 veces más que en la situación control. Estudios con construcciones artificiales donde un elemento potenciador se situaba a distintas distancias de un promotor (entre 75 -2500 pb) mostraron que el bucle activador entre el elemento distal y el promotor dependía de la unión de factores de transcripción específicos a ambos elementos (Nolis y cols 2009). Uno de los modelos para el estudio de los mecanismos moleculares que permiten la interacción de los elementos potenciadores distales con sus promotores diana ha sido la regulación de los loci de la α -globina y la β -globina. Mediante la aplicación de la técnica 3C (Captura de la Conformación de la Cromatina), una técnica que permite la detección y cuantificación de interacciones de larga distancia entre elementos de ADN (se verá en detalle en próximos apartados), se mostró como la eliminación de los factores de transcripción necesarios para la transcripción de la β -globina (EKLF y GATA-1), no sólo inactivaban el gen, sino que también, suprimían la interacción entre su promotor y su elemento cis-reguladora “región de control” (LCR, “locus control region”) (Drissen y cols 2004, Vakoc y cols 2005). En otros estudios se mostró como se producía una extrusión del locus de β -globina respecto de su territorio cromosómico circundante, específicamente en células proeritroblásticas, antes de la activación transcripcional de la β -globina. Esta extrusión dependía de la interacción con el LCR, y probablemente mantenía al locus de β -globina en un estado de “preparación”, indicando que la interacción del LCR con el promotor de la β -globina no es una consecuencia de la activación sino que la precede, y posiblemente la causa (Ragoczy y cols 2003). Por otro lado la integración ectópica del LCR puede provocar la extrusión de la región circundante y los genes que ésta contenga respecto de su territorio cromosómico, pudiendo llegar a producir la sobreexpresión de algunos de ellos (Noordermeer y cols 2008). Mediante el direccionamiento artificial de co-factores (Ldb1) al promotor de la β -globina, en células KO para GATA-1 (TF esencial para la interacción entre el LCR y el promotor de la β -globina), se conseguía rescatar el bucle activador entre el LCR

y la β -globina, recuperándose la transcripción de la misma. Estudios a escala global, muestran que existe una correlación entre la interacción con un mayor número de potenciadores y la mayor actividad transcripcional de un promotor (Chepelev y cols 2012). Esta serie de indicios pone de manifiesto que los potenciadores juegan un papel activo en la regulación de los genes a través de la reorganización estructural de la cromatina. Teniendo en cuenta la ingente cantidad de posibles potenciadores en comparación con el número de genes, y que no siempre los potenciadores se asocian a los genes más cercanos en la secuencia lineal (Lettice y cols 2002, Smemo y cols 2014), identificar los genes dianas de dichos potenciadores y entender cómo se organiza a escala global la información reguladora, se ha convertido en uno de los principales objetivos de la biología del desarrollo.

Desde un punto de vista de la aproximación experimental podríamos decir que existen dos tipos de métodos para asociar un determinado potenciador con su o sus promotores diana. El primer tipo de métodos serían aquellos métodos físicos que directa o indirectamente proporcionan información sobre la interacción física o cercanía relativa entre distintos puntos de la cromatina. Esta información permite revelar la organización tridimensional de la cromatina y, en consecuencia, la asociación directa entre potenciadores y promotores. Entre estos métodos estarían los métodos basados en hibridación de sondas fluorescentes (FISH) y los métodos de Captura de Conformación de la Cromatina. El segundo tipo de métodos serían aquellos que, basándose en datos de la actividad de promotores y elementos reguladores (datos ARN-seq, datos de Chip-seq de modificaciones de histonas, datos de ADNaseI-seq, etc.) en distintos contextos (p.e distintos tipos celulares), infieren la asociación de los potenciadores con los promotores diana mediante la búsqueda de correlación en sus niveles de actividad en cada contexto. Estos dos tipos de métodos, lejos de ser exclusivos, se complementan entre sí. Así, mientras los primeros determinan una relación física directa entre los elementos, los segundos determinan una relación funcional. A continuación pasamos a ver más en profundidad la información que cada método es capaz de proporcionar.

2.3.1 Métodos Físicos

a) Hibridación in situ fluorescente - FISH.

La hibridación in situ fluorescente (FISH) es una técnica citogenética que permite el marcaje y visualización de regiones específicas del ADN mediante la hibridación específica de

secuencia con sondas complementarias sintetizadas y marcadas, directa o indirectamente, con fluorescencia. El uso de sondas estándar para FISH permite la detección de un locus particular, de un tipo de ADN o de cromosomas enteros (Volpi y Bridger 2008; Levsky y Singer 2003). Las sondas de ADN para FISH usadas para detectar pequeñas regiones del genoma se suelen generar a partir de librerías disponibles de BACs (Cromosomas Artificiales de Bacterias) o fósmidos. Para la detección de cromosomas completos o grandes regiones, las sondas se generan mediante PCRs de todas las secuencias de esa región. Además, en esta última década se ha desarrollado un FISH de alta resolución basado en oligonucleótidos marcados directamente, la cual permite una resolución al nivel de kilobases (Boyle y cols, 2011).

La aplicación del FISH y la microscopía de fluorescencia aplicada al estudio general de los cromosomas han revelado la existencia de territorios cromosómicos, tanto en células fijadas (Cremer y Cremer 2006) como en células vivas (Muller y cols 2010), los cuales se disponen de manera no aleatoria en el núcleo celular. La aplicación de esta técnica permitió observar como los cromosomas ricos en genes se agrupaban en el centro del núcleo mientras que los cromosomas con menos concentración génica se situaban en la periferia nuclear (Croft y cols 1999, Boyle y cols 2001, Kupper y cols 2007, Kölbl y cols 2012). Esta tecnología reveló a su vez como la cromatina activa se descondensaba y separaba del resto del cromosoma, y permitió observar como las regiones de cromatina activa tendían a agruparse entre sí, incluso con regiones activas de diferentes cromosomas (Mahy y cols 2002; Müller y cols 2010; Boyle y cols 2011; Brown y cols 2008). A escala de sub-megabase, la aplicación del FISH ha revelado diferencias en la compactación de la cromatina interfásica en distintas regiones de una célula, de una misma región a lo largo de un proceso de diferenciación o de una misma región en una célula WT comparada con una mutante (Yokota y cols 1997; Morey y cols 2007). Finalmente, cabe destacar que la aplicación del FISH de alta resolución se ha utilizado para el estudio de la colocalización específica de tejido de elementos cis-reguladores y sus promotores diana (Amano y cols 2009; Williamson y cols 2012.)

b) Métodos de Captura de la Conformación de la Cromatina (3C)

El entorno o paisaje regulador de un gen se podría definir como el conjunto de la cromatina que alberga el total de la información cis reguladora de dicho gen. En definitiva, el conjunto de cromatina que contiene todos los elementos cis-reguladores que intervienen en la correcta expresión de dicho gen. El tamaño y complejidad de este paisaje regulador, por tanto,

tenderá a ser proporcional a la complejidad transcripcional de dicho gen. Así encontramos que los genes de desarrollo, que orquestan multitud de procesos diferentes en lugares diferentes durante el desarrollo embrionario, están asociados a grandes paisajes reguladores con multitud de elementos funcionales en cis. Las técnicas de captura de la conformación de la cromatina, han permitido estudiar el grado o frecuencia de interacción entre distintas regiones de la cromatina (Gavrilov y cols 2009). Los elementos reguladores de un promotor en particular, tenderán a estar localizados en regiones de la cromatina que posean una frecuencia de interacción significativa (respecto del fondo) con dicho promotor. Esto ha hecho que las técnicas de captura de conformación de la cromatina se hayan convertido en una herramienta fundamental en el estudio y determinación de los paisajes reguladores de los genes. En 2002 Dekker y cols desarrollaron la tecnología de Captura de la Conformación de la Cromatina (3C). Esta técnica bioquímica permite analizar la frecuencia de contactos entre dos regiones del genoma en una población celular. Desde entonces, se han desarrollado métodos genómicos derivados del 3C que, en comparación con métodos basados en la microscopía, permiten un análisis sistemático de la organización del ADN a una resolución mucho mayor. Los mapas o matrices de contacto (representaciones gráficas de la frecuencia de interacción entre distintos fragmentos de ADN a lo largo de un locus o genoma completo) nos ayudan a entender las reglas que determinan la organización y funcionamiento del genoma en el interior de la célula. Es importante tener en cuenta que las conclusiones que se derivan de estos métodos son probabilistas y no deterministas. Esto es, no se puede inferir la organización o funcionamiento del genoma de una célula individual, sino un comportamiento medio en una población celular. A continuación pasamos a describir y comparar los distintos métodos derivados de la técnica de la Captura de la Conformación de la Cromatina (Fig. I_3).

b).1 3C: Captura de la Conformación de la Cromatina.

La esencia de esta técnica y sus derivados es obtener una representación real de la organización tridimensional del genoma en el interior celular. Para ello, el primer paso es la fijación de las interacciones proteína-proteína y proteína-ADN mediante el uso de agentes fijadores, generalmente paraformaldehído. De esta forma se consigue una foto fija de las interacciones entre fragmentos de ADN que están mediados por proteínas. El siguiente paso consiste en la digestión con una enzima de restricción específica, la frecuencia de corte de esta enzima contribuirá a la resolución final de la técnica (p.e: DpnII reconoce una secuencia

de 4 pb y cortará con mayor frecuencia que HindIII que reconoce una secuencia de 6 pb). El siguiente paso consiste en una dilución de la muestra y una religación de los fragmentos por sus extremos cohesivos, lo que promueve la ligación intramolecular (esto es, entre los fragmentos que se encuentran fijados entre sí por proteínas). De esta forma, fragmentos lejanos en la secuencia lineal del ADN, pero que interactuaban entre sí mediante proteínas, se encuentran ahora fusionados en una misma molécula de ADN.

La forma de establecer la organización espacial de un locus particular, se conseguiría midiendo el número de eventos de ligación entre los distintos fragmentos del locus. En el caso del 3C, esta medición se consigue mediante PCRs cuantitativas o semicuantitativas. Para ello se diseñan cebadores cerca y hacia el extremo de cada fragmento de interés. Comparando la eficiencia en la amplificación de cada combinación de cebadores, se obtiene una matriz de frecuencia de ligación de la que se puede inferir la frecuencia de interacción entre los distintos fragmentos.

Esta técnica ha servido para demostrar la existencia de bucles de ADN o interacciones entre elementos reguladores y sus promotores diana tales como: la región “LCR” y el gen de la β -globina (Tolhuis y cols 2002), la Región de Control Global (“GCR”) y el gen *Hoxd13* (Montavon y cols 2011), o bucles implicados en la regulación del complejo IrxA (Tena y cols 2011). También ha permitido dilucidar que la acción aisladora de ciertos elementos de ADN es consecuencia del establecimiento de bucles u organizaciones tridimensionales específicas (Splinter y cols 2006, Zhao y cols 2006). Además permitió el descubrimiento de que interacciones promotor-terminador de un mismo gen incrementaban su capacidad transcripcional, como es el caso de ARN ribosómico (Nemeth y cols 2008).

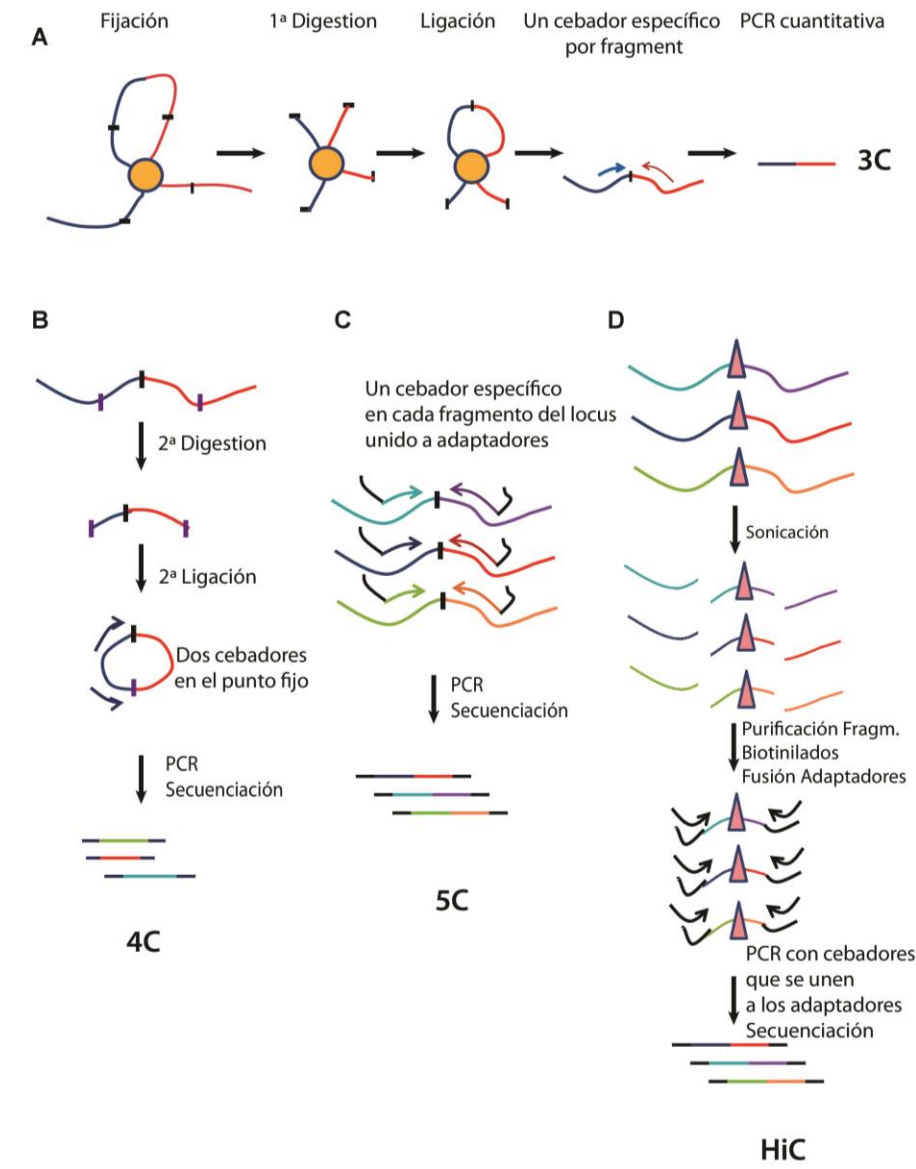
A la hora de interpretar las frecuencias de contactos obtenidas a partir de la técnica de 3C y sus derivados es necesario tener en cuenta que es lo que realmente se está “capturando”. Lo que estas técnicas nos muestran es la frecuencia con que, en una población celular, dos puntos del genoma se encuentran en proximidad física. Sin embargo, estas técnicas no distinguen si existe o no relevancia funcional en la cercanía de esos dos puntos o que mecanismos hay detrás de esa co-localización. La proximidad espacial puede ser el resultado de un contacto específico y directo entre dos puntos mediado por un complejo proteico que los une, o un resultado indirecto de la co-localización de una región genómica en una subestructura nuclear específica (p.e: lámina nuclear, nucléolo o factoría de transcripción). Además la colocalización espacial entre dos puntos puede ser debida indirectamente al

resultado de la organización tridimensional de una región de la cromatina promovida por otros contactos distintos, o por colisiones aleatorias en el núcleo. Este tipo de consideraciones han de tenerse en cuenta a la hora de interpretar el significado biológico de los datos generados por esta técnica.

b).2. 4C- Captura Circular de la Conformación de la Cromatina

Esta versión permite obtener la frecuencia de interacción que tiene un fragmento específico del genoma con respecto al resto de fragmentos del genoma. Se entiende como estrategia “uno vs todo”. El protocolo técnico incluye una variación con respecto al 3C (ver detalles en la sección de Material y Métodos). En resumen, la muestra generada con un 3C es digerida de nuevo con una enzima distinta, y religada otra vez, de manera que permite formar pequeñas moléculas circulares de ADN compuestas por pares de fragmentos capturados. Mediante PCR inversa con cebadores específicos del fragmento de interés, se amplifican todos los fragmentos que interaccionan con él. Cada fragmento se amplificará con una intensidad que dependerá de la frecuencia de interacción con el punto de interés. Finalmente, mediante secuenciación masiva de la librería derivada de la PCR inversa, se obtiene un mapa genómico de la frecuencia relativa de interacción de cada punto del genoma con tu fragmento de interés (Simonis y cols 2006, Fullwood y cols 2009).

Esta técnica se ha usado, por ejemplo, para comparar el perfil de contactos del promotor de un gen con expresión específica de tejido (β -globina) con el perfil de un gen ubicuo (Rad23), mostrando que el perfil del primero variaba en función de su estado regulatorio mientras que el perfil de segundo se mantenía constante (Simonis y cols 2006). También ha sido fundamental su uso para dilucidar los mecanismos moleculares que hay tras la co-linealidad en la expresión de los genes *Hox* (Montavon y cols 2011, Noordermeer y cols 2011). En general la técnica de 4C está concebida para obtener el perfil de contactos a nivel genómico de una región particular del ADN. Aplicada a un promotor, sirve para obtener el posible paisaje regulador del mismo; mientras que aplicada a un potenciador, nos ayudaría a determinar cuáles son los potenciales genes diana del mismo.



3 Fig I_3

Fig. I_3: Métodos de Captura de la Conformación de la Cromatina: Se representan las principales etapas en la generación de muestras para el estudio de la conformación de la cromatina. **A) Etapas comunes.** El principio de esta técnica es la fijación de las interacciones proteína-proteína y proteína-ADN. Luego mediante la digestión con una enzima de restricción particular y un proceso de ligación, se fusionan en una sola molécula los pares de fragmentos de ADN genómicos que estuvieran interaccionando entre sí. Esta moléculas fusionadas se pueden detectar de forma individual mediante PCR cuantitativa diseñando un primer en cada fragmento de la interacción (**3C "uno vs uno"**). **B) 4C ("uno vs todo")**. Esta variante requiere una nueva etapa de digestión con una enzima de restricción diferente y una nueva ligación. De esta forma se circularizan las moléculas formadas por pares de fragmentos interaccionantes. Una PCR con dos cebadores diseñados en un mismo fragmento pero orientados hacia fuera, amplificaría todos los fragmentos de ADN que hubiesen estado interaccionando con nuestro fragmento de interés. **C) 5C ("muchos vs mucho")**. Sin necesidad de procesar más la muestra se diseña un primer en cada fragmento de un locus particular y se realizan todas las combinaciones de PCR posibles. **D) HiC ("todo vs todo")**. Previa a la primera ligación se rellenan los extremos de los fragmentos generados en la digestión con nucleótidos biotinilados. Tras la primera etapa se sónica la muestra y se seleccionan pequeños fragmentos de los cuales se purifican los biotinilados y se les añaden los adaptadores para la secuenciación.

b).3. 5C – Copia de Carbón de la Captura de la Conformación de la Cromatina.

Si al 4C se le considera una estrategia “uno vs todo”, se podría decir que por el contrario el 5C es una estrategia “muchos vs muchos”, ya que permite obtener el perfil de frecuencia de contacto de todos los fragmentos a lo largo de una región determinada. En el 5C, la muestra obtenida en el 3C se hibrida con una mezcla de oligonucleótidos, compuesta por cebadores que se superponen parcialmente a los diferentes sitios de restricción de la región genómica que quieras estudiar. Los pares de cebadores que correspondan a dos fragmentos del genoma capturados entre sí, al ser hibridados sobre la muestra de 3C se yuxtapondrán de forma que pueden ser ligados en una misma molécula. Los oligonucleótidos llevan acoplados en la región 5' secuencias universales que permiten la amplificación simultánea por PCR de todos los pares de cebadores ligados entre sí. Esta PCR puede someterse luego a secuenciación masiva, obteniéndose una señal para cada par específico de cebadores proporcional a la frecuencia con la que han sido ligados entre sí. La resolución de la técnica viene determinada por la distancia entre los cebadores diseñados a lo largo de la región genómica estudiada., pero nunca podrá ser tan alta como la del 4C, porque no todos los sitios de restricción permiten el diseño de cebadores adecuados.

Esta tecnología permite la identificación de interacciones elemento cis-regulador-promotor a escala semigenómica. De hecho la aplicación de esta técnica sobre el locus de la β -globina (Dostie y cols 2006), o de los clústeres HOXA-D en humanos (Fraser y cols 2009; Ferraiuolo y cols 2010; Wang y cols 2011), ha confirmado interacciones previamente descubiertas con 3C así como nuevas interacciones con elementos reguladores hasta entonces desconocidas. La gran ventaja es, sin embargo, que el 5C genera una matriz de frecuencias de contactos entre todos los puntos de una región del genoma. Esta matriz sirve de semilla y punto de partida para generar la estructura tridimensional que dicha región adquiere.

b).4. HiC

El HiC es la versión “todo vs todo” del 3C, ya que genera una matriz de frecuencia de ligación de todos los fragmentos del genoma entre sí. Esta técnica fue originalmente desarrollada por Lieberman-Aiden y cols en 2009. El HiC requiere una pequeña

modificación en la generación de la muestra del 3C. Básicamente después de la primera digestión, los extremos cohesivos de los fragmentos se rellenan con nucleótidos biotinilados. Tras este paso se produce una dilución de la muestra y una ligación de los extremos romos. La muestra se des-fija y se purifica el ADN, se sónica y realiza una precipitación de biotina, de forma que se asegura que sólo los fragmentos ligados pasan a secuenciarse. Cuando la secuencia de un fragmento ligado contiene regiones de dos puntos distintos del genoma, se considera que es una interacción de estos dos puntos. La resolución de esta técnica tiene una dependencia cuadrática con la profundidad de la secuenciación, esto es con la cantidad de “reads” que se proporcionen. Así en 2009 Lieberman-Aiden y cols obtuvieron una resolución de 1 Mb a partir de 10 millones de “reads”, mientras que con 4,9 billones de “reads” se consiguió una resolución de 1 kb (Rao y cols 2014). La aplicación de esta técnica al genoma de distintos mamíferos ha permitido comenzar a entender los principios organizativos del genoma interfásico.

2.3.2 Métodos basados en la correlación de actividad

Estos métodos se basan en la integración mediante herramientas computacionales de multitud de datos ómicos relacionados con la actividad de promotores y potenciadores tipo RNA-seq, Chip-seq, ADNaseI-seq, para predecir la asociación entre potenciadores y sus promotores diana. Un ejemplo de este tipo de métodos ha sido el llevado cabo por Ernst J y cols 2012. En este trabajo se generaron datos de Chip-seq para 8 marcas distintas de modificación de histonas y para sitios de unión de la proteína CTCF, en 9 tipos celulares distintos para la caracterización sistemática de elementos reguladores y su especificidad celular. Esto se integró con datos de expresión génica y de predicción de sitios de unión para factores de transcripción. En resumen, para la asociación de un potenciador a un promotor diana, se buscaba en las 50 kb aguas arriba y aguas abajo, que promotor tenía un perfil de actividad y un patrón de sitios de unión de factores de transcripción parecido al del potenciador. Esto, sin embargo, introduce un sesgo, ya que limita a 50kb la distancia a la cual un potenciador puede localizarse respecto de su promotor diana. En otro estudio (Zen y cols 2012) se generaron datos de Chip-seq de H3K4me1 (para el estudio de potenciadores) y de polII (para el estudio de promotores) en 19 tejidos y cultivos celulares distintos de ratón. Para asignar potenciadores a sus promotores diana se calculó el coeficiente de correlación de

Spearman (SSC) para cada uno de los posibles pares de potenciadores-promotores, asociando aquellos con la mayor correlación a lo largo de las 19 muestras. Al contrario que en el caso anterior, este método no tiene ningún sesgo basado en la distancia. Este estudio mostró como los promotores y potenciadores co-regulados tendían a localizarse agrupados, y que la extensión de estas agrupaciones coincidían con regiones con alta frecuencia de interacción física (este último dato se obtenía de los datos derivados de la aplicación de HiC). Otro método de asociación de potenciadores y promotores se basó en la detección a escala genómica de sitios hipersensibles a la acción de la ADNasaI (DHSs) a lo largo de 79 células o tejidos humanos distintos (Thurman y cols 2012). Esto permitió el establecimiento de pares potenciadores-promotores de alta correlación. Todos estos métodos de asociación por correlación, deben ser validados con métodos físicos, principalmente con métodos de Captura de la Conformación de la Cromatina. Sin embargo, estos métodos de asociación presentan algunas limitaciones o sesgos. La primera, se relaciona al hecho de que dos elementos se activen con un mismo perfil a lo largo de distintos contextos, lo que puede ser reflejo de que compartan la regulación aguas arriba pero no, necesariamente, que uno de los elementos sea diana del otro. Por otro lado, un promotor puede ser diana de multitud de potenciadores, por lo que no necesariamente tiene que correlacionar su actividad con la de cada uno de sus potenciadores de forma individual.

3. ORGANIZACIÓN TOPOLÓGICA DEL GENOMA Y SU IMPLICACIÓN EN LA REGULACIÓN GÉNICA

Una de las características más llamativas del genoma eucariota, es su enorme tamaño lineal cuando se compara con las dimensiones del núcleo celular que lo alberga. Esto implica que el genoma debe sufrir un empaquetamiento extraordinario para poder ser alojado en el interior nuclear. Además esto debe ocurrir al mismo tiempo que millones de regiones específicas deben de estar accesibles a la enorme y compleja maquinaria proteica encargada de la transcripción de miles de genes en respuesta a multitud de estímulos diferentes. Esta exigencia y versatilidad funcional del inmenso genoma de los metazoos, restringida en el limitado dominio espacial del núcleo, ha suscitado siempre el interés por determinar si existe una organización topológica específicamente diseñada para responder a tales requerimientos funcionales.

3.2 DOMINIOS TOPOLÓGICAMENTE ASOCIADOS (TADs)

La aplicación a escala genómica de la tecnología de Captura de la Conformación de la Cromatina y sus derivados, ha revelado que el genoma de muchos metazoos se organiza en módulos estructurales de en torno a 500kb – 1Mb llamados dominios físicos o con mayor frecuencia dominios topológicamente asociados (TADs, del inglés (Topologically Associating Domains)(Dixon y cols 2012, Nora y cols 2012, Sexton y cols 2012) (Fig. I_4, A). Los contactos entre regiones genómicas que pertenecen a un mismo TAD se producen con mucha mayor frecuencia que los contactos entre regiones que pertenecen a TADs adyacentes. Los TADs transcripcionalmente activos tienden a asociarse preferentemente con otros TADs activos, mientras que los silenciados tienden a asociarse con otros silenciados (Lieberman-Aiden y cols 2009). La importancia de estos TADs es que no sólo actúan como unidades de organización estructural, sino que parecen también actuar como unidades funcionales. Así los TADs correlacionan con muchas marcas de actividad de la cromatina tales como: modificaciones de histonas (Dixon y cols 2012, Sexton y cols 2012), expresión génica coordinada (Le Dily y cols 2014, Nora y cols 2012), asociación a la lámina nuclear (Dixon y cols 2012) y coordinación de la replicación del ADN (Pope y cols 2014). A pesar de todos los estudios llevados a cabo en los últimos años, no se ha podido aclarar la cuestión de si la organización tridimensional del genoma en estos dominios estructurales es una causa o una consecuencia de las funciones genómicas (Cavalli y Misteli 2013, de Laat y Duboule 2013).

En realidad se conoce poco sobre los mecanismos moleculares que establecen y mantienen estos dominios estructurales. Dichos dominios físicos parecen estar regulados de forma autónoma unos de otros, y limitar el rango de acción de los elementos reguladores distales (Ghavi-Helm y cols 2014; Shen y cols 2014; Symmons y cols 2014). La comparación a escala global de los TADs en diversos tipos celulares de ratón y humano muestran que, en la mayoría de los casos, los límites de los TADs permanecen invariables entre distintos tipos celulares (Dixon y cols 2012). Esta invariabilidad, no se produce sólo en los dominios que contienen desiertos génicos o genes de expresión ubicua, sino también en dominios que contienen genes cuyo estado regulatorio cambia según el tipo celular. Esto indicaría, que al menos la configuración global y los límites de la mayoría de dominios topológicos, no son

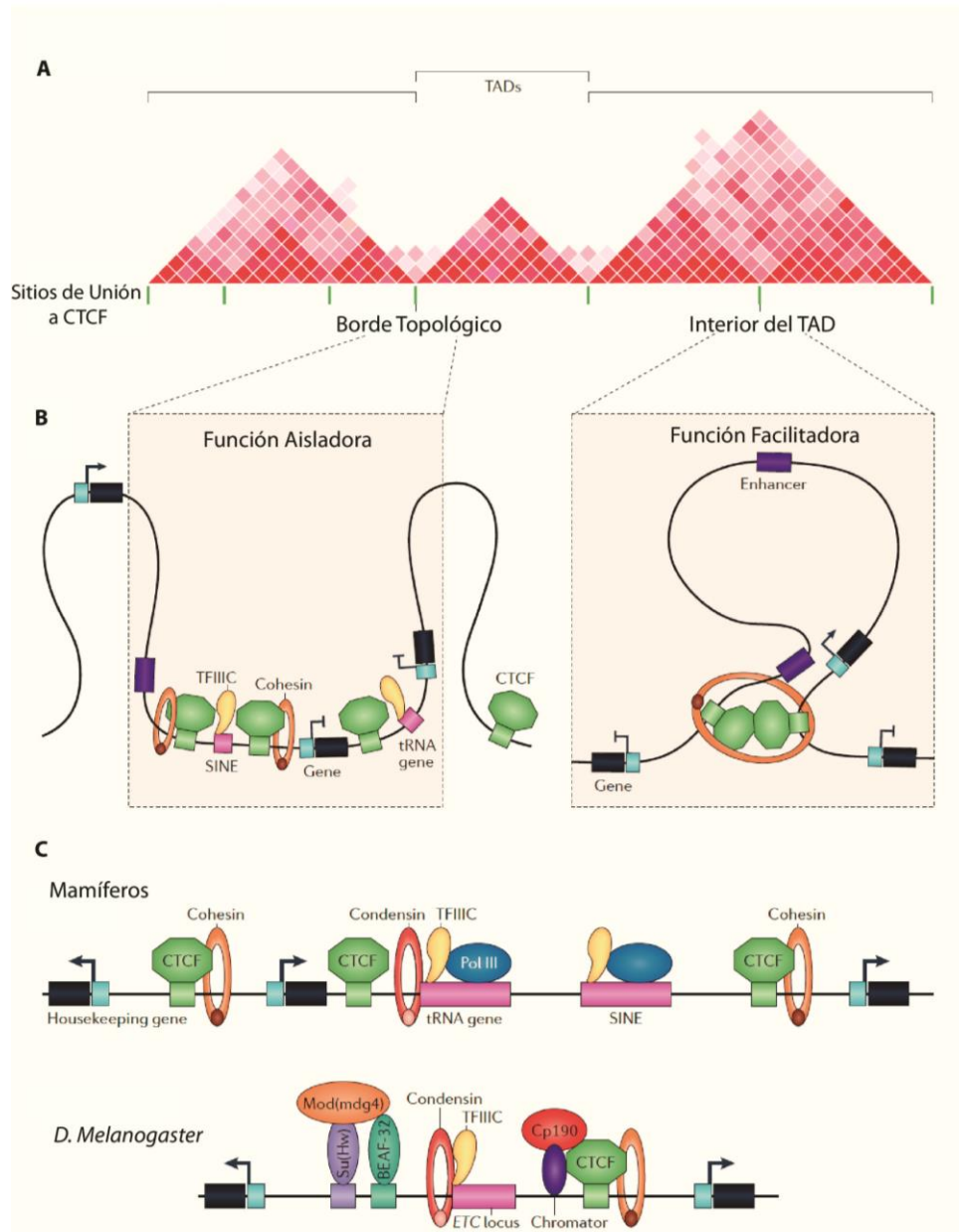
consecuencia de la actividad transcripcional de los genes que albergan. Por tanto se podría considerar que la mayor parte de los TADs son constitutivos. Por otro lado, el estudio particular de la colinealidad espacial y temporal de los genes *Hox* en ratón revela un comportamiento distinto para algunos dominios topológicos concretos (Andrey y cols 2013; Noordemeer y cols 2014; Lonfat y Duboule 2015). Los genes *Hox* se expresan secuencialmente tanto en el tiempo como en el espacio (eje anteroposterior o durante el desarrollo de la extremidades en el eje próximo distal) en un orden que viene determinado por su posición relativa en la secuencia genómica lineal. Los genes *Hox* que pertenecen a un mismo clúster se encuentran repartidos entre dos TADs adyacentes (al menos en el caso de los vertebrados estudiados). Los genes activos están marcados por H3K4me3 mientras que los inactivos lo están por H3K27me3 (modificaciones postraduccionales de histonas asociadas con actividad o silenciamiento transcripcional respectivamente). El proceso secuencial de activación de estos genes va acompañado de una transición en la modificación de las histonas asociadas a los genes, pasando de H3K27me3 (inactiva) a H3K4me3 (activa). En este caso el borde entre los dos TADs que se reparten los genes *Hox* de un mismo clúster, va variando posicionando los genes activos a un lado y los inactivos al otro. Estos dominios físicos, cuyos límites varían en función del estado transcripcional de los genes que contienen, se considerarían como dominios topológicos “facultativos”. Quizás en gran medida la estabilidad y variabilidad de los TADs dependa en muchas ocasiones de la resolución de la “lupa” con la que analizamos las interacciones a nivel global. El análisis con alta-resolución (mediante el uso de 5C en vez de HiC) de varios TADs durante el proceso de diferenciación de células madre embrionarias de ratón, muestra como a pesar de que algunos TADs permanecen predominantemente estables durante el proceso de desarrollo, se producen variaciones organizativas internas (Phillips-Cremins y cols 2013). Estas regiones que dentro de los TADs varían su conformación en función del estado regulatorio de la célula, se las denomina “sub-TADs”.

3.3 *PROTEÍNAS ARQUITECTÓNICAS, INTERACCIONES CONSTITUTIVAS Y BORDES TOPOLÓGICOS.*

No sólo los factores de transcripción específicos de tejidos son capaces de mediar la interacción física entre dos regiones de ADN distanciadas en la secuencia cromosómica lineal. Existen una serie de proteínas ubicuas que han sido asociadas con el establecimiento

de contactos cromosómicos de larga distancia. Este grupo de proteínas ubicuas reciben el nombre de “proteínas arquitectónicas” por su contribución a la organización tridimensional del genoma interfásico. Desde levaduras hasta humanos encontramos proteínas que ejercen o son susceptibles de ejercer este tipo de función “arquitectónica” (Ong y Corces 2014). Así por ejemplo encontramos el factor asociado a la ARN-polimerasa III, TFIIC, o la condensina y la cohesina (Heck 1997), descubiertas originalmente por su papel en la compactación y segregación de los cromosomas, y que ahora sabemos que tienen un papel muy relevante en el establecimiento de contactos funcionales entre elementos reguladores y genes diana (Merckenslager y Odom 2013; Faure y cols 2012). El complejo co-activador de la ARN-Polimerasa II, *Mediator*, también se ha involucrado en el establecimiento de estos contactos cromosómicos funcionales de larga distancia (Schaaf y cols 2013). Este complejo co-activador serviría de puente entre la maquinaria general de la transcripción presente en los promotores y los factores de transcripción específicos localizados en los elementos distales, estabilizando dicha conexión mediante su interacción con el complejo *cohesina* (Kagey y cols 2010). El complejo *Mediator* puede influir en la conformación de la cromatina mediante otros mecanismos, tales como la interacción con ARNs no codificantes o la modificación de histonas (Lai y cols 2013, Ding y cols 2008).

La “proteína arquitectónica” mejor caracterizada es la proteína de unión al sitio CCCTC, más conocida como CTCF. Esta proteína no está solo presente en mamíferos, sino también en otros vertebrados como el pez cebra (Pugacheva y cols 2006) y la rana (Burke y cols 2002) así como en muchos invertebrados, tales como la mosca de la fruta (Moon y cols 2005) y el mosquito (Gray y Coates 2005). De hecho se ha postulado que la proteína CTCF ha jugado un papel fundamental en la diversificación del phylum de los organismos bilaterales durante el Cámbrico (Heger y cols 2012). Esta proteína fue inicialmente descrita como un represor transcripcional (Lobanenko y cols 1990), aunque posteriormente también se caracterizó como un activador transcripcional (Klenova y cols 1993, Filippova y cols 1998). Quizás la función más característica asociada a CTCF ha sido la de aislador transcripcional (Bell y cols 1999). Cuando CTCF se posiciona entre un potenciador y su promotor diana, puede impedir la comunicación física entre ambos y en consecuencia la activación transcripcional del promotor. De hecho, en algunos casos se ha mostrado que esta función aisladora se realiza mediante la formación de bucles de ADN que alejan el elemento regulador del promotor (Kurukuti y cols 2006). En contraposición, muchos sitios de unión a CTCF no



Modificada de Ong y Corces 2014

4 Fig I_4

Fig. I_4: Organización del genoma en Dominios Topológicamente Asociados y el papel de CTCF.

A) Representación esquemática de datos de HiC obtenidos en células de mamíferos. A mayor intensidad del rojo de un cuadrado mayor es la frecuencia de contacto entre los 2 fragmentos correspondientes a la posición indicada por la proyección sobre el eje X de las diagonales que pasan por el cuadrado. Los TADs y los sitios de unión a CTCF se encuentran indicados. **B)** La concentración, disposición o proteínas con las que interactúe CTCF en distintas posiciones del genoma ayudarán a determinar su función, que puede ser aisladora como en los bordes topológicos o facilitadora como en el interior de un TAD. **C)** Los bordes topológicos tanto de mamíferos como de *Drosophila* se encuentran enriquecidos en sitios de unión a proteínas estructurales de expresión ubicua, entre ellas CTCF.

muestran esta actividad aisladora (Sanyal y cols 2012) (Fig. I_4, B). La aplicación de técnicas de ChIP-seq con anticuerpos anti CTCF en diferentes tejidos de distintas especies (Schmidt y cols 2012, Wang y cols 2012), han mostrado que CTCF se une a miles de regiones a lo largo del genoma. Principalmente se une a regiones intergénicas frecuentemente alejadas de cualquier sitio de inicio de la transcripción. CTCF comparte sitios de unión con multitud de proteínas distintas, así encontramos que CTCF co-ocupa regiones de ADN con: la deacetilasa de histonas SIN3 (Lutz y cols 2000), con el receptor de la hormona tiroidea (Lutz y cols 2003) o del estrógeno (Ross-Ines y cols 2011), con la helicasa p68 (Yao y cols 2010), con la ARN-polimerasa II (Wada y cols 2009) y con otras “proteínas arquitectónicas” (Rubio y cols 2008, Kagey y cols 2010), entre otras (Fig. I_4, C). Al contrario que *cohesina*, CTCF no parece estar tan involucrado en interacciones de elementos reguladores con sus promotores sino, más bien, está asociado con interacciones constitutivas de larga distancia (Phillips-Cremins y cols 2013). Los sitios de unión a CTCF tienden a formar interacciones con otros sitios de unión a CTCF, siendo en la mayoría de los casos (>90%) una interacción entre CTCFs orientados de manera convergente (cabeza a cabeza) (Rao y cols 2014). Es particularmente interesante que a pesar de su distribución a lo largo de todo el genoma, existe un enriquecimiento de sitios de unión de CTCF en los bordes de regiones con distinta localización nuclear, en los bordes de regiones con distintas marcas epigenéticas o distinto estado transcripcional y en los bordes de los TADs (Guelen y cols 2008, Dixon y cols 2012, Hou y cols 2012). Si bien, la presencia exclusiva de CTCF no es suficiente para formar dichos bordes, ya que tan solo un 15% de los sitios de unión a CTCF en mamíferos se encuentran en los mismos (Dixon y cols 2012). Además, interacciones mediadas por CTCF han resultado ser fundamentales para el mantenimiento de dichos dominios topológicos (Giorgetti y cols 2014).

Así pues, en el genoma se combinan interacciones cromosómicas específicas que dependen del estado regulador de la célula y están mediadas por factores de transcripción específicos de tejido, con interacciones constitutivas independientes del tipo celular, mediadas principalmente por “proteínas arquitectónicas” ubicuas, que facilitarían y restringirían la extensión y capacidad de acción del primer tipo de interacciones. Es importante tener en cuenta que esta clasificación entre interacciones “facultativas” y “constitutivas” no es absoluta. Proteínas arquitectónicas y factores de transcripción pueden cooperar conjuntamente para facilitar tanto interacciones facultativas como constitutivas.

La conservación de la configuración global y los límites de los TADs en muchos tipos celulares, lleva a pensar que en gran medida estos dominios topológicos se definen en función de interacciones constitutivas. De hecho la comparación de los dominios topológicos en regiones sinténicas de distintos mamíferos, muestran una fuerte conservación en la organización y límites de dichos dominios que correlaciona con la conservación de los sitios de unión a la proteína CTCF (Vietri y cols 2015).

4. LOS GENES *Six*

El primer miembro de los genes *Six* en identificarse fue el gen *sine oculis* (*so*) de la mosca de la fruta *Drosophila melanogaster* durante un trabajo que recopilaba mutantes en la estructura, tamaño y pigmentación del ojo de esta mosca (Milani 1941). Mutaciones en el gen tenían efectos dramáticos en el ojo y el resto del sistema visual de la mosca. Posteriores trabajos identificaron otros 2 genes *Six*: *optix* y *Dsix4*. Mientras que *optix* se expresa también en la retina, *Dsix4* se expresa en derivados mesodérmicos tales como el músculo o las gónadas. El hecho de que otros genes involucrados en el desarrollo de la retina de *D. melanogaster* (p.e; *ey*, o *Pax6* en vertebrados) estuvieran también presentes en el desarrollo de la retina de vertebrados, fomentó la búsqueda de ortólogos de *so* y el resto de genes *Six*. Hoy en día se han encontrado miembros de los genes *Six* en al menos 50 vertebrados y 40 organismos invertebrados. Si bien la historia evolutiva de estos genes no está aún muy clara su origen se remonta, al menos, hasta antes de la separación de Bilaterales y Cnidarios hace unos 500 millones de años.

Los genes *Six* son una familia de factores de transcripción caracterizada por un dominio “homeobox” de unión a ADN y un dominio SIX, propio de esta familia, que permite la interacción con otras proteínas. En mamíferos encontramos 6 genes *Six* que, en base a su secuencia, pueden ser categorizados en 3 subfamilias: *Six1* y *Six2*; *Six3* y *Six6*; y *Six4* y *Six5* (Kawakami, 2000)(Fig I_5, A).

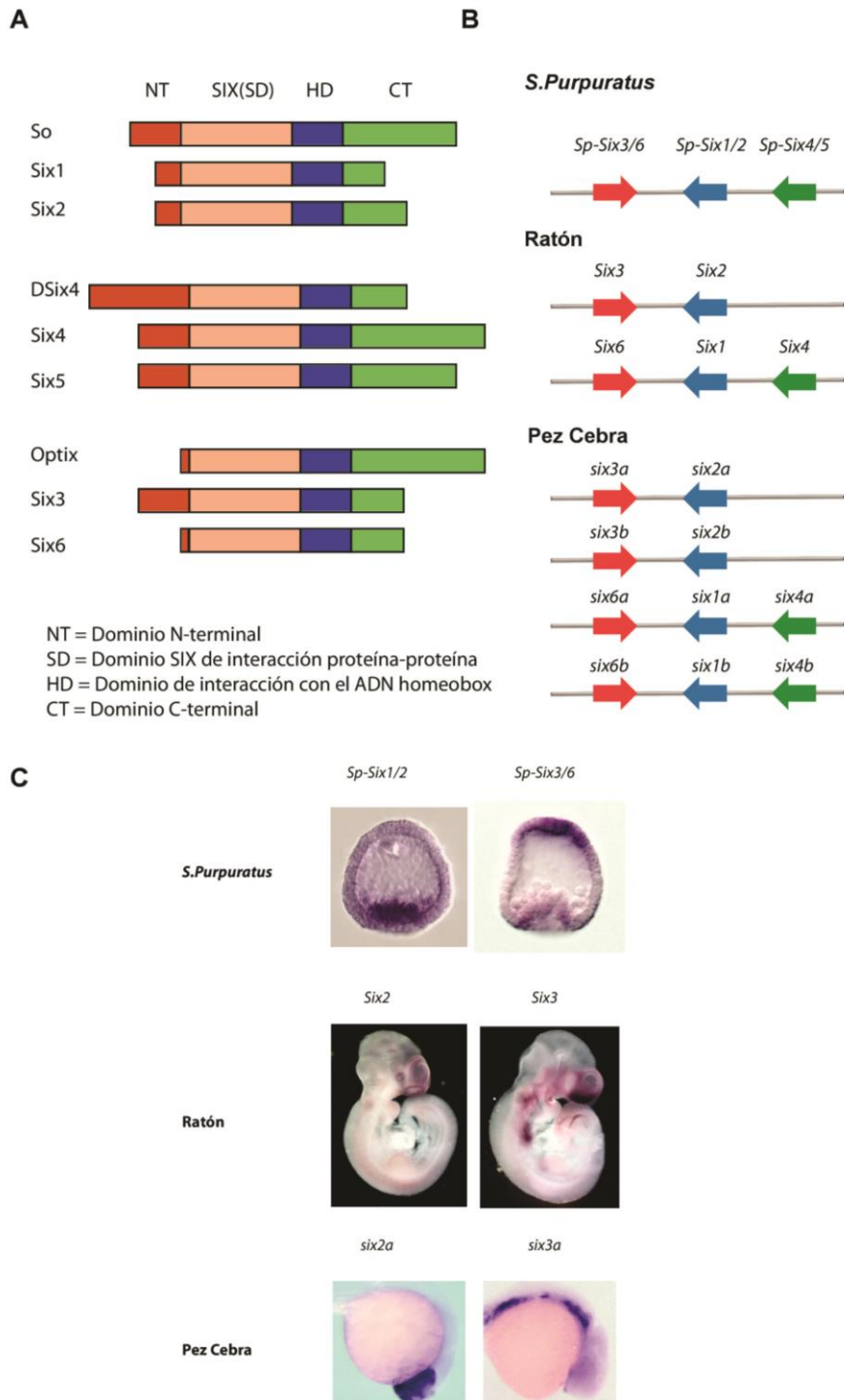
Evolutivamente se ha asociado la aparición de los genes *Six* con la aparición de la locomoción y de la predación. En general los genes *Six* intervienen fundamentalmente en el desarrollo de órganos sensoriales (retina, vesícula ótica, nariz) y de órganos locomotrices

(somitos, estructuras óseas, etc.). Estos genes están involucrados en procesos de proliferación celular, mantenimiento de un estado indiferenciado, supervivencia, migración e invasión, en aquellos tejidos donde son expresados (Ando H, 2005). Por ejemplo *Six3* se expresa principalmente en retina y prosencefalo. La sobre-expresión de *Six3* produce una expansión de esos tejidos, mientras que la inhibición produce una disminución (Ando y cols 2005; Gestri y cols 2005).

En ambos casos los tejidos se acaban formando y las células se diferencian correctamente, pero varía la cantidad de tejido, por lo que implica que *Six3* se encarga de mantener un stock determinado de células indiferenciadas y proliferativas (McCollum y cols 2007; Liu y cols 2010). Lo mismo ocurre con *Six2*, pero esta vez en estructuras óseas de nariz y mandíbula, o con el número de nefronas en el riñón (Fogelgren y cols 2008, Kobayashi y cols 2008,), tejidos donde *Six2* es expresado en vertebrados.

Precisamente en vertebrados estos genes se encuentran localizados principalmente en 2 complejos: *Six2-Six3* y *Six4-Six1-Six2* (Kawakami 2000). El mantenimiento de la sintenia de estos complejos se puede rastrear a lo largo de la evolución hasta al menos la base de los *Deuterostomos*, ya que en *S.purpuratus* (erizo de mar), por ejemplo, encontramos un único complejo *Six4/5-Six1/2-Six3/6* (Fig. I_5, B). Este complejo debió duplicarse en la duplicación genómica que ocurrió en la base de los vertebrados. Incluso la pareja *Six1/2-Six3/6* se encuentra ya en placozoos, considerados como los organismos base de los seres pluricelulares (Irimia y cols 2012).

El mantenimiento de bloques sinténicos de genes, como ocurre con los genes *Six*, se asocia al hecho de compartir elementos reguladores de larga distancia entre los genes del bloque, lo que impide que éstos se puedan alejar unos de otros, ya que perderían información reguladora fundamental. Esto es común en genes co-regulados o con patrones de expresión similares, o bien, en genes no relacionados evolutivamente donde uno de ellos alberga elementos reguladores del gen vecino manteniéndose junto a él,



5 Fig I_5

Fig. I_5: Los genes *Six*: dominios proteicos, organización genómica y expresión diferencial. A) En base a la similitud de los dominios proteico Homeobox y SIX, los genes *Six* se dividen en tres subfamilias: *Six3/6*, *Six1/2* y *Six4/5*. B) Organización recurrente de los genes *Six* en complejos génicos en Deuterostomos. Cada complejo está formado por miembros de distintas subfamilias de genes *Six*. C) Expresión diferencial de genes *Six* (*Six1/2* y *Six3/6*) de un mismo complejo en distintos organismos Deuterostomos.

incluso sin compartir dominios de expresión. El caso de los genes *Six* no se ubica en ninguno de las dos posibilidades anteriores, ya que tienen un origen evolutivo común, se han mantenido en sintenia y sin embargo tienen un patrón de expresión completamente distinto (Fig. I_5, C). El hecho de haber sido mantenidos físicamente ligados en el genoma, en muchas especies durante millones de años de evolución contrasta con el comportamiento transcripcional diferencial de los genes *Six* presentes en un mismo complejo. Cómo se sustentan unos paisajes reguladores tan próximos y ligados, y sin embargo tan diferentes, es la pregunta fundamental a la que pretende dar respuesta este trabajo. Además a través del estudio de esta situación particular, pretendemos contribuir al conocimiento general de los mecanismos moleculares responsables de la generación de paisajes reguladores y dominios topológicos. A través de la aplicación de técnicas de Captura de la Conformación de la Cromatina y de ensayos funcionales con BACs estudiaremos el papel de la organización tridimensional de los complejos *Six* en el desarrollo y la evolución.

Objetivos

Objetivos

Este trabajo quiere contribuir a entender cuál es el papel de la organización tridimensional de la cromatina en el control de la expresión génica, en particular la de los genes de desarrollo y durante el proceso de embriogénesis y la evolución. Para ello nos hemos centrado en una familia de genes de desarrollo que se organizan en complejos conservados en multitud de especies de Metazoos: los genes *Six*. A través del estudio de la organización 3D de la cromatina alrededor de estos complejos en pez cebra, ratón y erizo de mar pretendemos responder a:

- 1) ¿Existe un principio organizativo 3D de la cromatina común subyacente a los distintos complejos *Six*?
- 2) ¿Está relacionada esta organización 3D de la cromatina alrededor de los complejos *Six* con la forma relativa en que estos genes se expresan durante el desarrollo embrionario?
- 3) ¿Cuál es el impacto de la alteración de la organización 3D de la cromatina de los complejos *Six* en su regulación transcripcional durante la embriogénesis?
- 4) ¿Cuáles son los mecanismos moleculares responsables de dicha organización tridimensional?.

Materiales y Métodos

1. DETERMINACIÓN DEL ENTORNO REGULADOR DE UN GEN: 4C-SEQ.

Las técnicas de captura de la conformación de la cromatina o métodos “C” se fundamentan en la fijación de las interacciones proteína-proteína y ADN-proteína mediante el uso de paraformaldehído. Tras esta fijación, se realiza un paso de digestión con una enzima de restricción específica, se diluye la muestra y se realiza un paso de ligación. De esta manera dos fragmentos de ADN que estuviesen interaccionando quedarían fusionados en una sola molécula de ADN. En una muestra determinada, el número de fusiones entre dos fragmentos sería proporcional a la frecuencia con que esos fragmentos interaccionan en la población celular de tu muestra. La forma experimental para medir esa frecuencia de interacción es mediante PCR. Así si diseñas un cebador en el extremo de cada uno de esos dos fragmentos apuntando hacia la diana, obtendrás un producto de PCR que será proporcional al número de fusiones entre esos fragmentos.

En este trabajo, hemos usado la versión “Captura Circular de la Conformación de la Cromatina” o 4C-seq para determinar el paisaje regulador de los genes *Six* de pez cebra, ratón y erizo de mar. La técnica de 4C-seq permite identificar la frecuencia con que un fragmento de la cromatina interacciona con el resto de fragmentos del genoma en una muestra determinada, se trata de una técnica “uno vs todos”. Así, si centramos nuestro fragmento de estudio o “viewpoint” en el promotor de un gen, con la técnica de 4C-seq obtendremos la frecuencia con que este promotor interacciona con el resto de fragmentos del genoma. Esta información es crucial a la hora de delimitar el posible paisaje regulador de un gen. A continuación describimos el proceso de 4C-seq que hemos realizado usando muestras de pez cebra, ratón y erizo de mar.

1.1 *Preparación de muestras*

La preparación de las muestras se puede dividir en dos fases. La primera fase va desde la obtención de los embriones hasta la fijación de la muestra. Esta fase es específica para cada tipo de muestra o cada tipo de embrión. En nuestro caso, los embriones de cada especie presentan distintas particularidades que necesitan una adaptación del proceso, específica para cada uno de ellos. La segunda fase va desde el final de la fijación hasta la obtención de la muestra genómica de 4C final. Esta fase es común para todas las muestras. A continuación describimos la fase específica para cada tipo de muestra usada en este trabajo, seguida de la descripción de la segunda fase que es común para todas las muestras.

i. Primera Fase

-Pez cebra.

Se prepararon muestras para 4C-seq de embriones completos de pez cebra en distintos momentos del desarrollo. Para las muestras de 24hpf y 48hpf se utilizaron 500 embriones de pez cebra por muestra. Para las muestras de 80% de epibolia se usaron 1000 embriones por muestra, y para las muestras en el estadio de Dome se usaron 5000 embriones por muestra. Los embriones se obtienen mediante el cruce de individuos silvestres sexualmente maduros de la cepa AB, se recogen en medio E3 (NaCl 5 mM, KCl 0.17 mM, CaCl₂ 0.33 mM, and MgSO₄ 0.33 mM, pH 6.8–6.9), y se crecen a 28°C hasta alcanzar el estadio de desarrollo adecuado. Tras esto se elimina el corion de los embriones con un tratamiento con pronasa a una concentración de 1 mg/ml de E3, en el que los embriones se sumergen en esta solución y se agitan mecánicamente con una pipeta pasteur hasta que son decorionados. Los embriones decorionados se pasan a un tubo eppendorf junto con 1ml de la solución “*Ginsburg fish ringer*.” (NaCl 111 mM, KCl 3,6 mM, CaCl₂ 2,7mM, NaHCO₃ 1,9 mM), se pipetea con una pipeta P-200 con una punta amarilla hasta que los sacos vitelinos se disuelven. Tras esto se centrifuga el tubo eppendorf a 300 g durante 30 segundos, se retira el sobrenadante y se resuspenden los embriones en 10 ml de paraformaldehído al 2% en PBS en un tubo Falcón de 15 ml. Se mantiene a temperatura ambiente y en agitación durante 10 min. La fijación se para mediante la adición de 1,425 ml de glicina 2M, y la incubación en agitación a temperatura ambiente durante 5 min. A continuación se centrifuga 8’ a 1300 rpm a 4°C y se retira el sobrenadante. En este momento la muestra se puede guardar a -80°C previo paso por nitrógeno líquido o bien continuar con la lisis.

-Ratón.

Las muestras de ratón se obtuvieron siguiendo protocolos ya descritos (Hagège y cols 2007; Dekker y cols 2002; Splinter y cols 2012). Embriones completos de ratón a 9.5 y 14.5 dpf respectivamente fueron procesados para obtener diversas muestras de 10 millones de células cada una. El procesamiento consistió en la extracción de los embriones en el momento de gestación adecuado, la disgregación de los mismos mediante técnicas mecánicas (tijeras, pinzas, etc.) un posterior tratamiento con colagenasa durante 45 min a 37°C, y por último un homogenizado y tamizado hasta obtener una suspensión de células tamizadas. El número de células obtenidas se calcula mediante el uso de una Cámara de Neubauer. El total de células se divide en grupos de 10 millones en tubos Falcón de 15ml si es necesario usando PBS como solución. Se realiza un pequeña centrifugación para precipitar las células y retirar el PBS y se resuspenden en 10 ml de paraformaldehído al 2% en PBS. Al igual que con las muestras de pez cebra se mantienen a temperatura ambiente y en agitación durante 10 min. La fijación se para mediante la adición de 1,425 ml de glicina 2M, y la incubación en agitación a temperatura ambiente durante 5 min. A continuación se centrifuga 8’ a 1300 rpm

a 4°C y se retira el sobrenadante. En este momento la muestra se puede guardar a -80°C previo paso por nitrógeno líquido o bien continuar con la lisis.

-Erizo de Mar.

Para cada muestra de 4C de *S. purpuratus* entre 10000 y 100000 embriones a 48 hpf (~1000 células por embrión). Siempre se trabaja en una habitación a 15 grados. Individuos adultos fueron inyectados con KCl 0,5M en el interior del celoma y agitados suavemente para la extracción de los gametos. Los óvulos son recogidos en agua de mar filtrada mientras que el esperma se recoge de la superficie del erizo de mar con una pipeta pasteur y se almacena en un tubo eppendorf. Tratamos de partir de 1×10^6 embriones que dividimos en 10 partes para preparar 10 muestras (las condiciones de fijación para las muestras de erizo de mar no están optimizadas, y esto nos permite modificar algunas variables (tiempo, temperatura y % de PFA) y probar). Para la obtención del número de embriones necesarios fecundamos normalmente el contenido de óvulos total de una o varias hembras. Para ello colocamos los óvulos en un vaso de precipitado con agua de mar filtrada y añadimos unas gotas de esperma. Removemos suavemente y esperamos unos minutos. Tomando una pequeña muestra de los óvulos, comprobamos al microscopio el porcentaje de óvulos fecundados. Para considerar que la fecundación ha funcionado correctamente debemos obtener un porcentaje de fecundación cercano al 100%. Si no es así se añade más esperma y se repite esta acción varias veces. Si aun así no conseguimos una correcta fecundación se recomienda obtener nuevos óvulos y nuevo esperma. Una vez comprobada la fecundación, crecemos los embriones en 5 litros de agua de mar filtrada con agitación a 15°C hasta alcanzar el estadio de desarrollo adecuado. Es importante comprobar cada cierto tiempo el correcto desarrollo de los embriones. Una vez alcanzado el estadio adecuado los embriones se concentran en 10ml en un tubo falcón de 15 ml mediante sucesivas decantaciones. Se comprueba el estado de los embriones al microscopio y se hace una estimación del número de embriones final mediante diluciones y conteos al microscopio. Tenemos que dejar la muestra a una concentración de ~100000 embriones por ml, y dividirla en tantos eppendorf como ml tengamos. De esta forma obtenemos muestras individuales de 100000 embriones. Como hemos mencionado, el proceso de preparación de las muestras de erizo de mar para 4C no está optimizado. Hemos generado multitud de muestras en condiciones diferentes, para ello hemos manejado las variables: proceso de celularización embrionaria, tiempo de fijación, temperatura de fijación y porcentaje de paraformaldehído usado en la fijación. Sin embargo, no ha sido posible por el momento comprobar cómo funciona cada una de las condiciones. Así que a continuación nos limitamos a describir el proceso de fijación usado para los 4C que se presentan en este trabajo.

Trabajamos a 4°C. Los 100000 embriones de una muestra son disociados en búfer de disociación enfriado en hielo (glicina 1M pH8, EDTA 2mM, y un cóctel con inhibidores de proteasas 1x). Ayudar al proceso de celularización con una pipeta. Comprobar este proceso al microscopio. Una vez terminado lavar varias veces con agua de mar libre de calcio (CaFSW) (NaCl 437mM, KCl 9mM, MgCl₂ 6 H₂O 22,9 mM, MgSO₄ 7 H₂O 25,5 mM,

NaHCO₃ 2,1mM). Este paso es crítico ya que para que la fijación funcione debemos eliminar todos los restos de glicina. Recomendando al menos seis pasos de lavado con 1ml de CaFSW. En cada paso de lavado perderemos muestra, pero 100000 embriones de partida es una cantidad más que suficiente que permite la pérdida de material en el proceso. A continuación lavamos una vez en PBS. A continuación se fijan las muestras en un tubo falcón de 15 ml con 10 ml de paraformaldehído al 2% en PBS durante 10 min a temperatura ambiente en agitación. La reacción de fijación se para añadiendo glicina a una concentración final de 125 mM durante 5 minutos (temperatura ambiente y agitación). Tras parar la fijación las células se centrifugan 8' a 1300 rpm a 4°C y se retira el sobrenadante. Como el resto de muestras, en este punto la células pueden continuar con el proceso de lisis o ser guardadas a -80°C pasando por nitrógeno líquido.

ii. Segunda Fase

Una vez fijadas, todas las muestras siguen el mismo proceso independientemente de su procedencia. Comienza aquí la segunda fase de preparación de la muestra de 4C. El siguiente paso es pues la lisis. Para este paso resuspendemos las células en buffer de lisis (50mM Tris-HCl pH 7.5, 150mM NaCl, 5mM EDTA, 1% TX-100 (n° cat. T8787, Sigma Aldrich), 0,5 % IGEPAL CA-630 (n° cat I8896 Sigma-Aldrich;), coctel inhibidor de proteasas 1X (CØmplete, n° cat 11697498001, Roche Applied Science)) enfriado en hielo. Y pasamos las células a un homogeneizador Dounce de vidrio situado en hielo. Este tipo de homogeneizadores permiten la rotura de las células manteniendo, en gran medida, la integridad del núcleo. Si bien la mayoría de la bibliografía describe el proceso de lisis como un paso de 5 min, en nuestras manos suele alargarse más incluso con el uso del homogeneizador, no siendo extraño llegar a los 30 min. La eficiencia de la lisis se comprueba cada 5-10 min cogiendo 3µl de la misma y mezclándolos con 3µl de la solución de tinción Methyl Green- Pyronin (n° cat. HT70116, Sigma-Aldrich) que marca el citoplasma en rosa y el núcleo de las células en verde. Una buena lisis se correspondería con una muestra donde la mayoría de citoplasmas se ha disgregado quedando los núcleos verdes intactos. Si bien nunca hemos obtenido una lisis perfecta. En determinados casos, para favorecer la lisis es necesario aumentar la concentración de IGEPAL CA-630 en el buffer entre dos y tres veces. Una vez finalizada la lisis se centrifugan las células a 4°C 5 min, se elimina el sobrenadante y se resuspenden en 450 µl de agua doblemente destilada. El resto del protocolo detallado se puede consultar en el ANEXO 1. En resumen, el resto del proceso consiste en una primera digestión con la enzima de restricción DpnII. Después de comprobar la digestión la enzima se inactiva por temperatura y la muestra se diluye y se somete a un proceso de ligación mediante la enzima T4 ligasa. Esto permite, al estar fijadas las interacciones, fusionar en una sola molécula de ADN dos fragmentos (DpnII-DpnII) que estuviesen interaccionando. Una vez comprobada la ligación, la muestra se “des fija” mediante un tratamiento con proteinasa K y alta temperatura y se precipita. Se realiza una nueva ronda de digestión con la enzima Csp6I y otra ligación con T4 ligasa. De esta forma

conseguimos pequeños círculos de ADN formados por pares de fragmentos (DpnII-Csp6I) que estaban interaccionando en el núcleo vía proteínas.

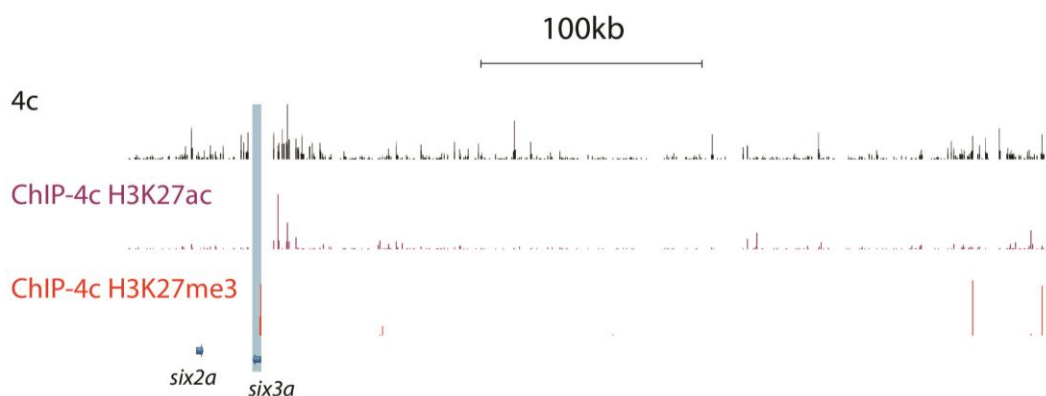
iii. Chip – 4C

En un intento de enriquecer nuestras muestras en contactos funcionales respecto de contactos estructurales, hemos llevado a cabo una modificación del protocolo general de 4C-seq, al que hemos llamado con el nombre de ChIP-4C. Este procedimiento experimental está lejos de haber sido validado. Se trata tan solo de una “prueba de principio”, una aproximación rápida y no sistemática para tratar de arrojar luz sobre una cuestión concreta. Ningún resultado derivado de este procedimiento puede conducir a ningún tipo de conclusión, por tanto, los comentarios que en este trabajo se realizan deben enmarcarse siempre en el campo de la especulación razonada.

La idea es que incluso en un contexto experimental muy acotado, por ejemplo un tipo celular muy específico donde nuestro gen se encuentra muy activo transcripcionalmente, las interacciones estructurales pueden enmascarar a las funcionales, dada su naturaleza más estable. Si pudiésemos separar específicamente las interacciones funcionales de las estructurales dentro de una misma muestra, podríamos obtener un perfil de contactos genuinamente funcionales. Para ello nos servimos de “marcas funcionales” asociadas a la cromatina. En nuestro caso nos hemos fijado en dos. Una “marca funcional” asociada a activación transcripcional, la modificación de histona H3K27ac. Y una “marca funcional” asociada a silenciamiento transcripcional, la modificación de histona H3K27me3. Si podemos enriquecer nuestras muestras en fragmentos asociados a las marcas H3K27ac y H3K27me3, estaremos enriqueciendo la muestra en contactos funcionales asociados específicamente con la activación o silenciamiento transcripcional, separándolos de los contactos estructurales que son más frecuentes y siempre están presentes.

Para llevar a cabo este enriquecimiento selectivo de las muestras tratamos de llevar a cabo un paso de inmunoprecipitación con anticuerpos contra H3K27ac y H3K27me3, intercalado en el procesamiento normal de una muestra de 4C. Este proceso de inmunoprecipitación lo llevamos a cabo tras el paso de la primera ligación y antes del tratamiento con proteinasa K. La modificación detallada del proceso se puede consultar en el ANEXO 2. Tras esta modificación el proceso sigue de forma estándar pero sin comprobar en gel el paso de la segunda digestión, debido a que la cantidad de muestra que obtenemos tras la precipitación es muy limitada.

Precisamente la cantidad de muestra es el mayor problema de esta aproximación. En este caso hemos partido del material estándar de un 4C por lo que la cantidad final de muestra es muy baja. Esta cantidad funciona para ChIP-seq pero es muy baja para 4C, dando lugar a perfiles de 4C muy sesgados (Fig. MyM_1). Se podría escalar la cantidad de muestra pero requeriría escalar la cantidad de anticuerpo, lo que haría el proceso demasiado costoso.



6 Fig. MyM_1

Fig. MyM_1: Comparación de los perfiles de contacto obtenidos por 4C y por ChIP-4C. Para un mismo punto fijo (barra azul) se muestran las “lecturas mapeadas” obtenidas tras realizar un 4C estándar, un ChIP-4C para H3K27ac y un ChIP-4C para H3K27me3 sobre embriones completos de pez cebra a 24 hpf. Se observa un perfil mucho más discontinuo para los ChIP-4C que para el 4C estándar.

1.2 Diseño de cebadores

El diseño de cebadores es muy importante en esta técnica y hay que tener en cuenta varios criterios a la hora de diseñarlos. Lo primero es elegir el punto fijo o “viewpoint”. Hay que considerar primero el criterio biológico: si pretendemos saber cómo interacciona el promotor de un gen, nuestro punto fijo debe contener a dicho promotor o estar lo más cerca posible del mismo. Lo segundo son los criterios técnicos. Debemos buscar, en nuestro caso, un fragmento delimitado a un lado por la enzima DpnII y al otro por la enzima Csp6I. Este fragmento debe tener al menos un tamaño de 150 pb para que se facilite la recircularización en los pasos de ligación. Además se deben de evitar los fragmentos con secuencias repetitivas y de baja complejidad para facilitar el diseño de cebadores en su interior.

Para el diseño de cebadores, como en otras ocasiones, se usó el programa en línea Primer3 (Rozen y Skaletsky. 2000) con los parámetros siguientes: longitud del cebador: 18-27, óptimo 20; temperatura de fusión (T_m): 58-62°C, óptimo 60°C; porcentaje de G/C: 40-60%; máximo poli-X: 3; terminación en G/C:2. Estos parámetros pueden variar en determinados casos en función de las necesidades. Particularmente para el diseño de cebadores para 4C hay que considerar los siguientes aspectos. En cada punto fijo, uno de los cebadores debe diseñarse en dirección a la diana de DpnII (cebador DpnII) y el otro hacia la diana de la enzima Csp6I (cebador Csp6) (Fig. MyM 2, A). El cebador DpnII, desde donde se realizará la secuenciación, debe diseñarse lo más de forma que su secuencia termine lo más cerca posible, sino encima, de la diana DpnII. Esto es crítico ya que la secuenciación posterior nos

dará “lecturas” de un tamaño de unos 36 pb. Si a estos 36 pares de bases le quitamos los alrededor de 20 pb correspondientes a la secuencia del cebador, sólo nos deja en torno a 16 pb para mapear dicha lectura en el genoma. Cada nucleótido que se separe el primer DpnII de la diana, será un nucleótido menos que no podrá ser usado para mapear la lectura. Si el cebador se separa demasiado de la diana corremos el riesgo de perder muchas de las “lecturas” que nos proporcione el secuenciador. El cebador Csp6I no es tan crítico, y puede diseñarse dentro de los 100pb adyacentes a la diana Csp6I. Los cebadores utilizados en este trabajo pueden encontrarse en el ANEXO 3.

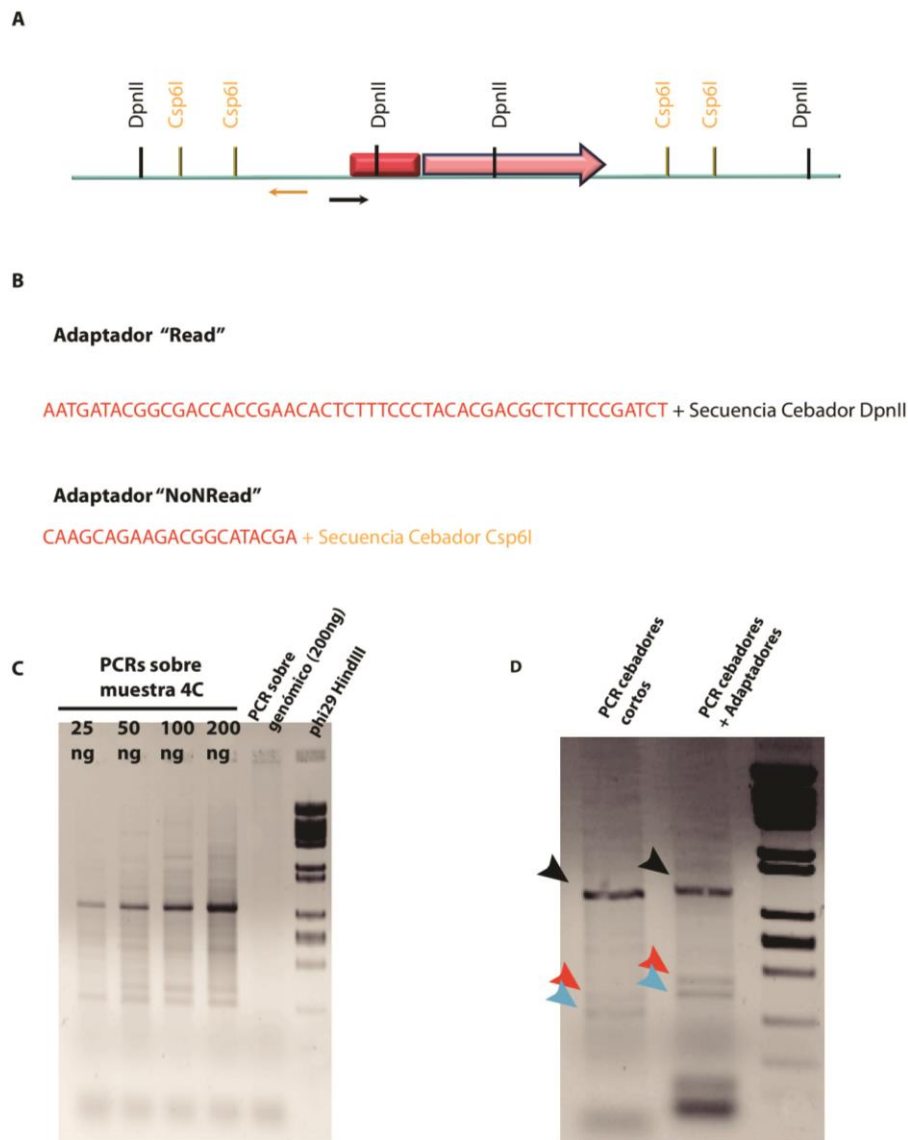
Otra cuestión a considerar respecto a los cebadores usados para 4C-seq es la siguiente. La idea es el uso de los cebadores para realizar una PCR sobre las muestras de 4C cuyo producto se someta a secuenciación masiva. Nosotros usamos secuenciación basada en la tecnología Illumina cuya descripción no compete a este trabajo y la cuál puede consultarse en la página www.Illumina.com. La cuestión es que para facilitar dicho proceso, las PCRs se hacen con nuestros cebadores fusionados a los adaptadores Illumina correspondientes (Fig. MyM 2, B). El cebador DpnII irá fusionado al adaptador “Read” y el cebador Csp6I al adaptador “NoNRead”. Esto hace que los cebadores alcancen un tamaño considerable por lo que al sintetizarse deben purificarse por métodos no estándar como HPLC o SDS-PAGE. Este tipo de purificación incrementa notablemente su precio. Por esta razón, previo a la compra de los cebadores fusionados a los adaptadores (“cebadores largos”), se prueban los cebadores DpnII y Csp6I sin fusión alguna (“cebadores cortos”) en PCRs sobre una muestra de 4C para comprobar su eficacia de amplificación.

1.3 PCRs y preparación de librerías de 4C

Una vez preparada la muestra, seleccionado el punto fijo y diseñados los cebadores, viene el proceso de obtención de la librería de 4C de nuestro punto de interés. Llamamos librería de 4C al conjunto de fragmentos generados por una PCR en rango lineal de amplificación, realizada con los cebadores DpnII y Csp6I fusionados a sus respectivos adaptadores de un punto fijo concreto y purificada. En este producto de PCR estarán representados los fragmentos genómicos que interaccionan con nuestro punto fijo en una cantidad relativa que será proporcional a la frecuencia con que su interacción se capturó en la muestra.

Para determinar el rango lineal de amplificación de nuestra PCR, realizaremos varias PCRs con cantidades crecientes de muestra. La prueba estándar consiste en cuatro PCRs con cantidades de 12,5, 25, 50 y 100 ng de muestra (estas cantidades pueden variar en función de la experiencia con determinadas muestras). La composición de la reacción de estas PCRs y las condiciones de los ciclos se puede consultar en el ANEXO 1. Estas PCRs se realizan con los “cebadores cortos” de forma que se comprueba la calidad de los mismos. Dado la diversidad y complejidad de tamaños de las moléculas de ADN que van a componer el producto de estas PCRs, es importante utilizar una polimerasa que no tenga (al menos en teoría) sesgo de amplificación por un rango concreto de tamaños. Para comprobar el

resultado de las PCRs, una parte de las mismas (15 µl de 25 µl totales), se separan en un gel de agarosa al 1,5 % (Fig. MyM 2, C). Si observamos



7 Fig MyM_2

Fig. MyM_2: Cebadores y PCRs para 4C-seq. **A)** Representación esquemática de una región a analizar mediante 4C-seq. El cebador DpnII (flecha negra) y el cebador Csp6I (flecha amarilla) se diseñan en un fragmento DpnII-DpnII que contiene al promotor (rectángulo rojo) del gen de interés (flecha roja). Este fragmento DpnII-DpnII debe contener en su interior una diana Csp6I. **B)** Adaptadores usados para la secuenciación con la plataforma Illumina. **C)** Puesta a punto de una PCR de 4C. Se prueban los cebadores cortos a distintas concentraciones para hallar el rango lineal. Podemos observar como la PCR amplifica multitud de fragmentos de distintos tamaños cuando usamos una muestra de 4C. Sin embargo, estos cebadores no amplifican nada cuando usamos genómico. **D)** Diferencia entre una PCR con los cebadores cortos y los fusionados a los adaptadores Illumina. Se observa un aumento en el tamaño de los fragmentos (flechas de colores).

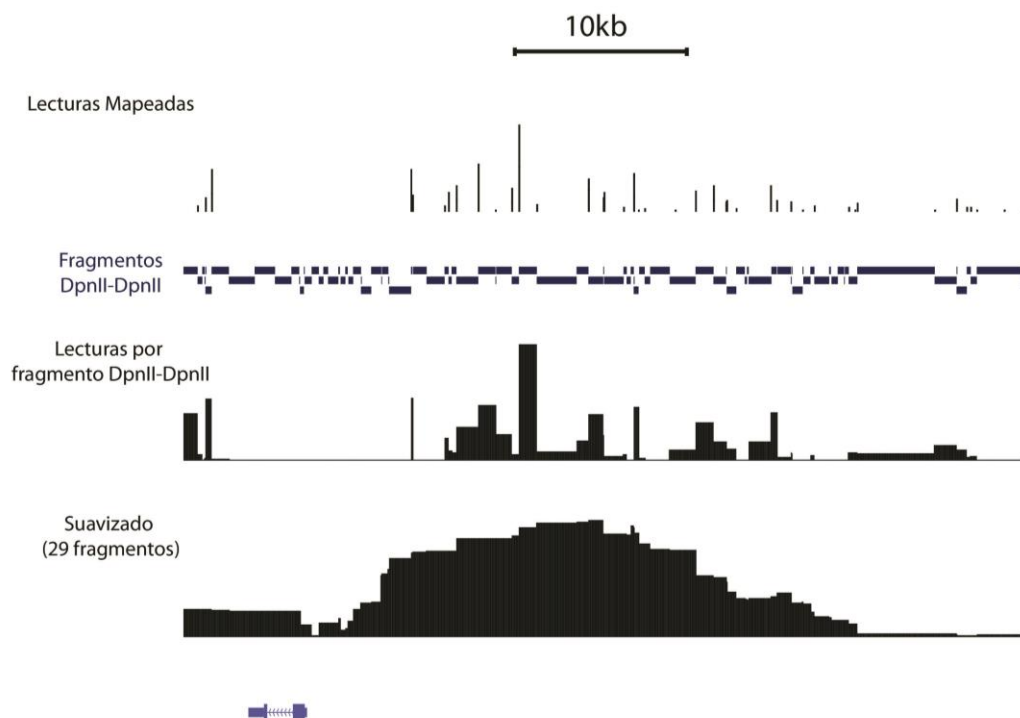
que el uso del doble de muestra produce aproximadamente el doble de producto de PCR, entonces estaremos en rango lineal de amplificación. Se determina la cantidad de muestra mínima que se necesita para obtener dicho rango lineal (el uso de la cantidad mínima es sólo un criterio de economía de muestra). Comprobada la eficiencia de los cebadores “cortos” y la cantidad de muestra necesaria, se realiza una prueba con los cebadores “largos” (fusionados a los adaptadores Illumina) en las condiciones determinadas. Deberemos observar un patrón de bandas similar al de los cebadores “cortos” pero con un desplazamiento de las bandas hacia arriba, debido al incremento del tamaño de los cebadores “largos” (Fig. MyM 2, D). Una vez que la prueba con los cebadores fusionados a los adaptadores es satisfactoria, realizamos la PCR que dará lugar a la librería de 4C en sí misma. La composición de esta PCR final se indica en el ANEXO 1. Una pequeña parte de esta PCR se separa en un gel de agarosa al 1,5% para comprobar que todo funcionó correctamente. El último paso es la purificación de la librería. Dado que el destino de la librería es la secuenciación masiva, la purificación es un paso crítico que tiene que ir encaminada a eliminar cualquier impureza que pueda interferir con el proceso de secuenciación. Además otra función importante de la purificación es eliminar los restos de cebadores, que podrían llegar a competir con los productos de PCRs en el proceso de secuenciación. Dada la importancia de este paso, nosotros realizamos dos purificaciones en tándem. Una primera purificación basada en el uso de columnas de fibra de vidrio (“High Pure PCR purification Kit”, nº cat 11732676001, Roche) seguido de una purificación con bolas magnéticas (“Agencourt AMPure XP PCR purification kit”, nº cat A63880, Beckman Coulter). Finalmente la concentración de la librería se mide mediante técnicas de tinción y detección fluorescente ultrasensible (“Quant-iT PicoGreen dsDNA reagents and kits”, cat nº P11496, Invitrogen).

En este punto una librería está lista para secuenciar. Una de las ventajas de la tecnología Illumina es la posibilidad de secuenciar conjuntamente en la misma ronda de secuenciación múltiples librerías. Las “lecturas” obtenidas en esta secuenciación pueden ser separadas por librerías atendiendo a la secuencia de los cebadores DpnII. El límite en el número de muestras que se pueden secuenciar conjuntamente debe ser explorado en cada caso. Nosotros secuenciamos conjuntamente en torno a 12-15 librerías de 4C sin perder calidad en los resultados.

1.4 Procesamiento de los datos

El procesamiento de los datos se realizó siguiendo los métodos estandarizados para esta técnica (Noordermeer y cols 2011, Smemo y cols 2014). Dado que secuenciamos conjuntamente varias librerías en la misma reacción de secuenciación, el primer paso consiste en la separación de los datos crudos por librerías. Esto se consigue teniendo en cuenta la secuencia del cebador DpnII. Todas la “lecturas” de una librería comenzarán con la secuencia correspondiente al cebador DpnII. Como esta secuencia es específica de cada punto fijo, basándonos en ella podemos separar todas las “lecturas” correspondientes a ese punto fijo. Para recuperar el máximo de “lecturas” por librería y teniendo en cuenta que

pueden ocurrir pequeños errores durante la secuenciación, podemos incluir cierto grado de ambigüedad (2 o 3 nucleótidos) entre la secuencia de



8 Fig MyM_3

Fig. MyM_3: Proceso de tratamiento de los datos de 4C. Se muestra el resultado de las diferentes etapas del procesado del producto de secuenciación de una librería de 4C. El panel superior muestra las “lecturas mapeadas” una vez eliminadas las correspondientes a las dos kilobases alrededor del punto fijo (en este caso el punto fijo es la secuencia codificante del panel inferior). El siguiente panel (en azul) muestra los distintos fragmentos DpnII-DpnII de la región. A continuación se muestra el resultado de extender las “lecturas mapeadas” a cada fragmento DpnII-DpnII completo, generando el perfil de “lecturas por fragmento DpnII-DpnII”. Por último se muestra el resultado del “suavizado” usando ventanas de 29 fragmentos DpnII-DpnII.

una “lectura” y la secuencia del cebador DpnII original. Por tanto los datos obtenidos en una secuenciación se pueden separar en diferentes librerías, mediante el uso de un script de Perl que busca la secuencia de cada cebador DpnII usado en todas las lecturas obtenidas y las separa en distintos archivos, en función del cebador al que correspondan. Una vez separadas las distintas librerías hay que mapear las lecturas sobre el genoma de referencia. En nuestro caso hemos usado la versión de Diciembre de 2009 para el pez cebra (Zv9/danRer7), la versión de Julio de 2007 de ratón (NCBI37/mm9) y la versión de Junio de 2011 para erizo de mar (BCM-HGSC Spur_v3.1/strPur4). Si bien no se tratan de las últimas versiones disponibles, son aquellas que han sido más usadas como referencia en la bibliografía y a las que hay asociadas un mayor número de datos experimentales. Para el mapeo de las lecturas presentes en cada librería hacemos uso del programa Bowtie. Este programa está especializado en el alineamiento ultra-rápido de enormes cantidades de fragmentos cortos de ADN sobre un genoma de referencia con un bajo uso de memoria, lo que permite el análisis

rápido de los datos, sin pérdida importante de precisión. Una vez mapeadas las “lecturas”, aquellas que han caído entre dos sitios de corte de la misma enzima de restricción o que han caído en fragmentos DpnII-Csp6I menores a 40 pares de bases son eliminados. A pesar de que una lectura sólo abarca en torno a los primeros 15 nucleótidos de un fragmento DpnII-DpnII, realmente representa la frecuencia de contacto de todo ese fragmento. Por este motivo, el siguiente paso tras mapear y filtrar las “lecturas” es pasar de “lecturas mapeadas” a “lecturas por fragmento DpnII-DpnII” (Fig. MyM_3). En este momento tenemos asignada una determinada “señal de contacto” (nº total de lecturas) a cada fragmento DpnII-DpnII. Dado que una PCR puede introducir sesgos para la amplificación de determinados fragmentos (en función del tamaño, porcentaje GC, etc.), a continuación se realiza un “suavizado” de los perfiles de contacto. Este “suavizado” consiste en coger una ventana de 30 fragmentos DpnII-DpnII, calcular la media de la señal de cada fragmento incluido en la ventana y poner el valor de señal resultante en el fragmento central de la ventana (Fig. MyM_3). A continuación mover esta ventana un fragmento y repetir el proceso. Así hasta recorrer el genoma completo. Estos perfiles de contacto suavizados son el resultado final del procesamiento de los datos de 4C, y representan la distribución de frecuencias relativa con las que un punto fijo concreto interacciona con el resto del genoma.

1.5 Cálculo del borde topológico predicho (btp).

Para poder comparar unos perfiles de 4C con otros, no podemos usar el número absoluto de “lecturas” ya que éste dependerá del proceso de secuenciación. Por tanto el primer paso es normalizar los datos. Para este fin, en cada complejo *Six* que estudiamos, seleccionamos una región que abarque la mayor parte de la señal de 4C de todos los genes *Six* presente en un complejo en particular. Una vez seleccionada la región a analizar en cada complejo y trabajando con los perfiles “suavizados”, calculamos el porcentaje de señal o porcentaje de “lecturas” que cada gen del complejo dedica a cada fragmento DpnII-DpnII. Así, en cada complejo, pasamos de una “señal absoluta” a una “señal relativa” que si podemos comparar entre distintos genes del mismo complejo.

Para predecir la posición de un hipotético borde topológico calculamos el gráfico del porcentaje acumulado de señal de cada gen de un complejo a lo largo de la región de estudio (Fig. RE_2, A). A continuación calculamos el punto donde la diferencia entre el porcentaje acumulado de señal entre dos genes de un mismo complejo es máxima. Además, mediante el cálculo de la diferencia de señal (sin acumular) de 4C entre genes supuestamente localizados en sitios opuestos del borde, estimamos si la región predicha se corresponde con el punto de transición en el que los contactos de un gen empiezan a disminuir considerablemente a la vez que los contactos del otro empiezan a aumentar bruscamente. Si efectivamente, ambos cálculos dan como resultado una región coincidente, consideraremos entonces que esa región es un hipotético borde topológico. Una vez establecido dicho borde, calculamos el porcentaje de señal que cada gen de un mismo complejo tiene a ambos lados del borde inferido.

Además basándonos en los perfiles “suavizados” y normalizados podemos analizar el grado de similitud entre distintos perfiles mediante el cálculo del coeficiente de correlación de Pearson (r).

2. SELECCIÓN DE POTENCIALES ELEMENTOS REGULADORES EN EL COMPLEJO *SIX2A-SIX3A*.

Para la caracterización funcional del entorno genómico del complejo *six2a-six3a* decidimos restringir la búsqueda de potenciales elementos cis-reguladores a la cromatina contenida en el BAC DKEY-74B2 (chr13:9,757,983-9,945,798). Este BAC de pez cebra abarca la región que va desde aproximadamente 120 kb aguas arriba de *six3a* hasta unas 50kb aguas abajo de *six2a*. Dicho BAC será posteriormente utilizado para el estudio funcional del posible borde topológico localizado entre ambos genes, y por tanto es de utilidad para caracterizar los elementos cis-reguladores contenidos en él. De esta forma podremos tener una visión más completa del impacto regulador de la desaparición del borde topológico sobre los genes *six3a* y *six2a*.

Para la selección de elementos reguladores en esta región tuvimos en cuenta tres criterios. El primero fue que se correspondiera a una región marcada con la modificación de histona asociada a elementos potenciadores activos: H3K27ac. Para ello contamos con los datos de ChIP-seq contra H3K27ac previamente generados en nuestro laboratorio (Bogdanovic y cols 2012) sobre embriones completos de pez cebra, en distintos momentos durante las primeras 48 hpf del desarrollo embrionario. El protocolo detallado para la realización de ChIP-seq sobre en embriones de pez cebra puede consultarse en Bogdanovic y cols 2013. El segundo criterio fue que el elemento se correspondiera a una región conservada en vertebrados. Para ello tuvimos en cuenta los datos de conservación del buscador genómico de la Universidad de California – Santa Cruz (<https://genome.ucsc.edu/>). Estos datos de conservación se miden usando dos métodos *phastCons* y *philoP*, y se calculan para todos los vertebrados y para un subconjunto de peces. El tercer y último criterio es que el elemento se encuentre en una región incluida en los perfiles de contacto de 4C de algunos de los genes (*six2a* y *six3a*).

Los cebadores utilizados para su amplificación y las regiones abarcadas por los mismos pueden consultarse en el ANEXO 4.

3. ENSAYO DE ELEMENTOS REGULADORES EN PEZ CEBRA

Los potenciales elementos reguladores seleccionados en el apartado anterior fueron testados mediante transgénesis en pez cebra para comprobar si albergaban o no algún tipo de actividad potenciadora de la transcripción. Se diseñaron cebadores específicos para las

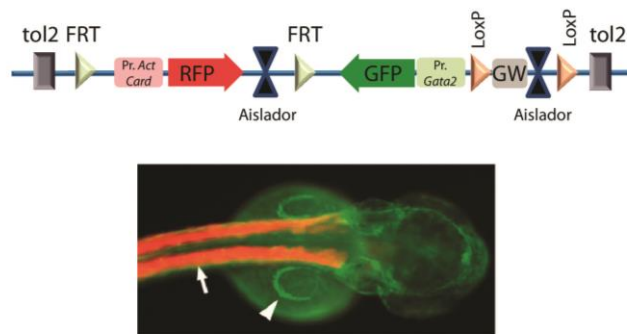
regiones seleccionadas (ANEXO 4), y éstas fueron amplificadas mediante PCR a partir de ADN genómico de pez cebra, utilizando para ello un termociclador *MyCycler* (BioRad). Los productos de PCR fueron corridos en un gel de agarosa (nº cat. 8014, Pronadisa) al 1% y purificados mediante centrifugación a través de una malla de acuario para separar el ADN de la agarosa. Posteriormente los productos de PCR fueron clonados en el vector pCR8/GW/TOPO usando las especificaciones del Kit de clonaje pCR8/GW/TOPO TA (Invitrogen, nº de catálogo K2500-20). A partir de este vector los elementos fueron recombinados en uno de los de plásmidos usados para testar actividad potenciadora mediante la transgénesis en pez cebra. La recombinación se realizó mediante el uso de la enzima clonasa (*Gateway LR clonase*, nº cat. 11791-019, Invitrogen). El método de transgénesis se basa en el uso de la transposasa Tol2 (Kawakami y cols 2000), que reconoce un par de secuencias específicas (L y R) siendo capaz de recombinar e integrar la región de ADN comprendida entre ambas en algún sitio aleatorio del genoma con una alta eficiencia. Los dos vectores utilizados para la transgénesis en pez cebra son:

- **Vector ZED (*Zebrafish Enhancer Detection vector*):** Publicado originalmente en Bessa y cols 2009 (Fig. MyM_4, A). En resumen, este vector contiene dos módulos contiguos que se encuentran flanqueados por las secuencias de reconocimiento de la transposasa Tol2. Un módulo está formado por el promotor de la actina cardiaca seguido del gen reportero *RFP* (“red fluorescent protein”) todo esto flanqueado por dos sitios FRT (estos sitios permiten, si fuese necesario, la escisión específica de este módulo mediante la acción de la enzima flipasa). Este módulo sirve de control interno de transgénesis pues independientemente de la actividad del elemento clonado siempre promoverá la expresión de *RFP* en el músculo del pez cebra. El segundo módulo consiste en un “*cassette*” de recombinación *Gateway* (flanqueado por dos sitios LoxP) seguido del promotor mínimo del gen *Gata2* y del gen reportero *GFP* (“green fluorescent protein”). Este módulo permite la recombinación de los fragmentos clonados en el vector pCR8/GW/TOPO TA justo delante del promotor mínimo de *Gata2*. Sólo en el caso de que el fragmento clonado tenga actividad potenciadora de la transcripción observaremos expresión de la *GFP*. A su vez este módulo se encuentra flanqueado por 2 secuencias aisladoras para evitar en la medida de lo posible interferencias reguladoras del entorno genómico donde el vector sea integrado.
- **Vector *JLz48*:** Este vector consiste en un cassette *Gateway* seguido del promotor mínimo del gen *Gata2* al que sigue el gen reportero *GFP* y una secuencia potenciadora de la transcripción (*z48*) que produce una expresión fuerte en el cerebro medio del pez cebra, todo ello flanqueado por las secuencias de reconocimiento de la transposasa Tol2 (Fig. MyM_4, B). Esta expresión en el cerebro medio constituye el control de transgénesis. Si el elemento clonado posee actividad potenciadora promoverá la expresión de *GFP* en territorios específicos que podremos identificar si son distintos del cerebro medio.

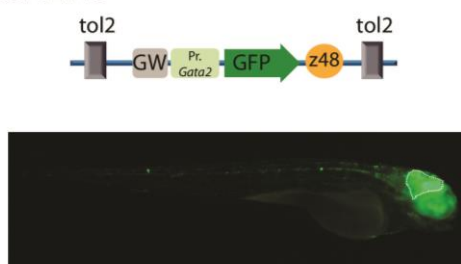
Una vez realizada la recombinación entre el vector donador (pCR8/GW/TOPO) y el vector de destino (ZED o JLz48) este último se purifica con el uso de columnas de purificación de plásmidos (GenElute Plasmid Miniprep Kit, n° cat PLN350, Sigma Aldrich o similar). Con el fin de eliminar cualquier endotoxina o impureza que pudiera disminuir la eficiencia en la transgénesis, esta construcción final fue sometida a un último proceso de fenolización (fenol:cloroformo 1:1) y precipitación con etanol absoluto, y su concentración fue cuantificada utilizando un espectrofotómetro (Nanodrop, Thermo Fisher Scientific).

Para realizar la transgénesis de los vectores en el pez cebra obtenemos una mezcla en agua libre de ARNasa consistente en: 25ng/μl del vector portador del elemento a testar, 25ng/μl del ARNm codificante para la transposasa Tol2 y 0,05% de rojo fenol (n° cat. P0290, Sigma-Aldrich). Alrededor de 2nl de esta mezcla se inyectan en embriones de una célula. Para ello se utilizó un capilar de vidrio de 1mm y 0,58mm de diámetro externo e interno respectivamente (n° cat. G100F6, Warner Instruments), estirado en un estirador horizontal programable (P-97 Flaming/Brown Micropipette Puller, Sutter Instrument) (IM-300, Narishige) mediante un micromanipulador de 3 ejes. A las 24 y 48 horas de la inyección, los embriones (entre 200 y 300 por construcción) son observados en una lupa fluorescencia (SZX16, Olympus). Aquellos embriones sospechosos de portar actividad reguladora y con un patrón homogéneo del respectivo control de transgénesis son crecidos hasta adultos (G0). Cuando estos embriones seleccionados alcanzan la madurez sexual (2,5 -3 meses) son cruzados con individuos WT. Para considerar a un elemento potenciador como positivo, necesitamos encontrar al menos tres adultos (G0) cuya progenie (G1) sea portadora del transgén y presenten unos patrones de GFP equivalentes (de forma que garanticen que dichos patrones no son producto de un efecto posicional).

A Vector ZED



B Vector JLz48



9 Fig MyM_4

Fig. MyM_4 Vectores para transgénesis en pez cebra. A) Vector ZED: El vector ZED se compone de dos bloques. Uno formado por el promotor de la *actina cardíaca* seguido del gen de la *RFP* que sirve como control interno de transgénesis y que se encuentra flanqueado por sitios FRT lo que permite su escisión mediante el uso de *flipasa*. El otro bloque consiste en un cassette Gateway para la recombinación de los elementos cis-reguladores a testar, seguido del promotor mínimo *Gata2* y del gen de la *GFP*. Este bloque sirve para probar la actividad potenciadora de las regiones clonadas y se encuentra flanqueado por dos aisladores fuertes. A su vez el cassette Gateway se encuentra rodeado por dos secuencias *LoxP* que permiten la escisión de la región clonada mediante el uso de la recombinasa *Cre*. Las secuencias de reconocimiento de la transposasa *tol2* flanquean ambos bloques **B) Vector JLz48:** Este vector se compone de un cassette Gateway para clonar las secuencias susceptibles de portar actividad potenciadora, seguido promotor mínimo *Gata2*, del gen de la *GFP* y por último de un potenciador fuerte (*z48*) que promueve la expresión en el cerebro medio y actúa como control

4. MODIFICACIÓN Y TRANSGÉNESIS DE BACS

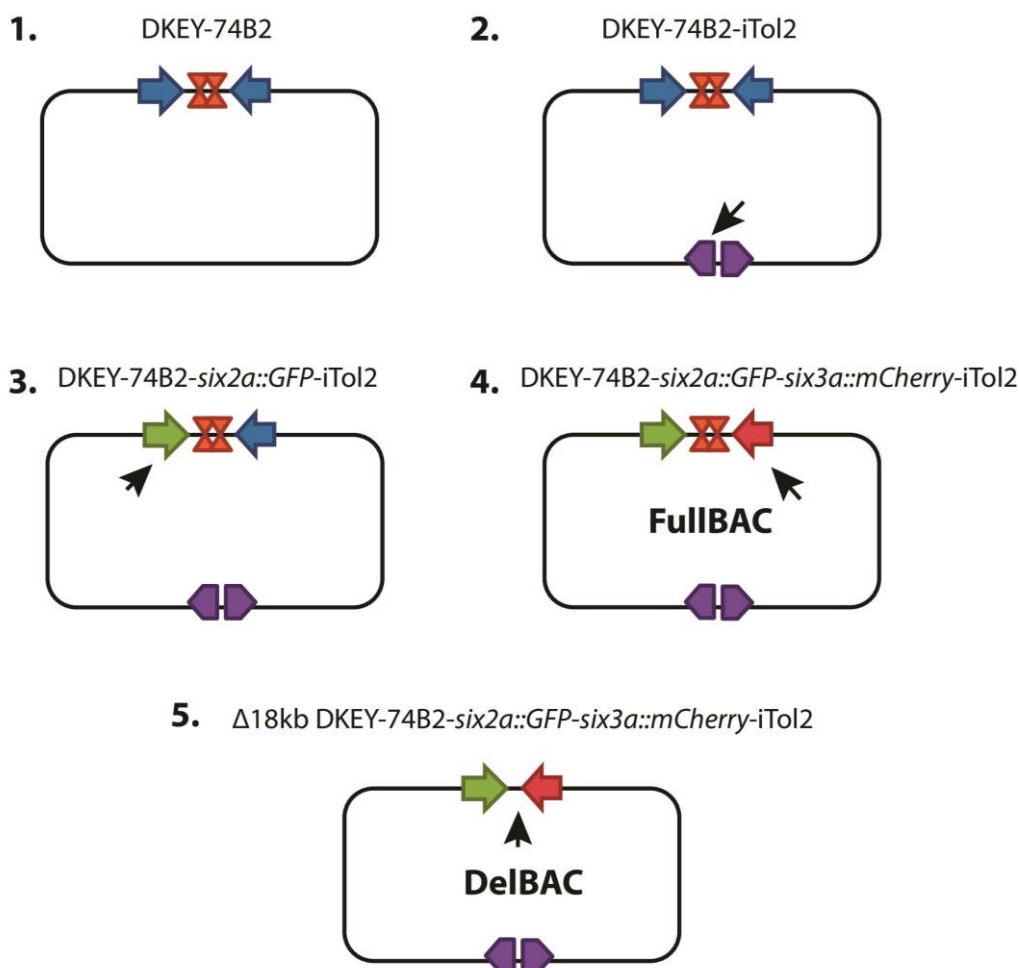
Un Cromosoma Artificial de Bacteria o BAC (del inglés Bacterial Artificial Chromosome) es un fragmento de ADN genómico, cuyo tamaño suele oscilar entre 100 y 300 kb, que se inserta en un vector que permite su replicación en cultivos bacterianos (Shizuya y cols 1992). Se utilizan para la generación de librerías y secuenciación de genomas (Choi y Kim 2001). Así mismo, el gran tamaño de estas moléculas de ADN les permite albergar gran cantidad de

información genética tanto codificante como reguladora. Por este motivo es común el uso de la transgénesis de BACs para el análisis del control de la expresión génica (Shakes y cols 2012, Torres-Núñez y cols 2015). Un único BAC es capaz, no sólo de contener la secuencia codificante completa o parcial de un gen o de varios, sino también una gran parte de la información reguladora de este gen.

Por otro lado, el gran potencial de la transgénesis de BACs para el estudio de la expresión génica radica en la posibilidad de modificar su secuencia (Ciotta y cols 2011; Narayanan y Chen 2011). Dado el gran tamaño de estas moléculas, su modificación mediante el uso de enzimas de restricción y ligasas resulta inapropiado (es muy común encontrar multitud de regiones diana para cualquier enzima de restricción). Por esta razón, supuso un gran avance el desarrollo de los métodos de modificación basados en la maquinaria de recombinación homóloga en bacterias, original de los fagos que la utilizan para insertar su genoma en la bacteria que infectaban (Murphy, 1998). Los métodos basados en este mecanismo se conocen como *recombineering*, acrónimo anglosajón derivado de “recombination-mediated genetic engineering” o en castellano “ingeniería genética mediada por recombinación”. Todos estos métodos se basan en la introducción de fragmentos de ADN flanqueados por regiones homólogas a las regiones diana del BAC, que se convertirán en el destino de la inserción. La maquinaria proteica necesaria para la recombinación homóloga puede ser introducida en la bacteria mediante un plásmido, aunque ya existen distintas estirpes de *E.coli* modificadas que presentan esta maquinaria en su genoma. Dicha maquinaria es inducible por *shock* térmico a 42°C. Tras esta inducción se electropora el ADN exógeno flanqueado por la regiones homólogas correspondientes, y se seleccionan aquellas bacterias con eventos positivos de recombinación. Aunque existen diversos métodos de selección, en este trabajo nos hemos basado en la selección por resistencia a antibióticos. El ADN exógeno que introducíamos en la bacteria iba acompañado de un cassette de resistencia a algún antibiótico, y tras su electroporación las bacterias eran cultivadas en medio con el antibiótico apropiado.

En este trabajo decidimos utilizar esta tecnología para el estudio del papel sobre la regulación transcripcional de los genes *six2a* y *six3a* del hipotético borde topológico presente entre ambos genes. Para ello seleccionamos un BAC (DKEYP-74B2) que abarca la región genómica que contiene a las secuencias codificantes y parte del entorno regulador de los genes *six2a* y *six3a*, incluyendo la región presente entre ambos genes donde se encontraría el borde topológico. La idea es sustituir la secuencia codificante de los genes *six2a* y *six3a* por genes reporteros (GFP y mCherry respectivamente). Tras esto, llevar a cabo la transgénesis en pez cebra de la versión completa del BAC y de una versión con el borde eliminado, y comparar ambos patrones de expresión (Fig. MyM_5).

Proceso de modificación por recombinación homóloga del BAC DKEY-74B2.



10 Fig MyM_5

Fig. MyM_5: Proceso de modificación por recombinación homóloga del BAC DKEY-74B2. El BAC DKEY-74B2 fue sometido a varias modificaciones en serie para dar lugar a los BACs FullBAC y $\Delta 18\text{kbBAC}$ listos para transgénesis. El primer paso fue la inclusión de los sitios inversos de reconocimiento de la transposasa tol2 en el esqueleto del BAC (2). El siguiente paso fue la inclusión de la GFP tras el primer ATG del gen *six2a* (3). Luego vino la inclusión del gen mCherry tras el primer ATG del gen *six3a* (4), lo que dio lugar a la versión FullBAC. Por último vino la eliminación de las 18kb correspondientes al hipotético borde topológico (4), dando lugar la versión $\Delta 18\text{kbBAC}$.

4.1 Generación del BAC *DK74B2-six2a::GFP-six3a::mCherry-iTol2*. (FullBAC)

El clone del BAC número 74B2 de la librerías de BAC DanioKey se obtuvo de la empresa Source BioScience. Se realizaron PCRs sobre diversas regiones y varias digestiones (XbaI, EcoRI, BglII) para comprobar la identidad e integridad del mismo. La versión DK74B2-*six2a::GFP-six3a::mCherry-iTol2* se generó siguiendo métodos descritos anteriormente en la bibliografía (Bussmann y cols 2011, Suster y cols 2011). El BAC se purificó usando el kit

Nucleobond BAC 100 (Macherey-Nagel) y se electroporó en la estirpe de *E. coli* SW105 (Warming y cols 2005). SW105 es una estirpe que deriva de la DY380 que contiene al profago λ defectivo con los genes de necesarios para la recombinación homóloga controlados por un represor sensible a la temperatura. Además SW105 está diseñado con sistema de selección *galK* (no se usa en este trabajo) y contiene el gen *flpe* controlado por un sistema inducible por arabinosa. El producto proteico de *flpe* media la recombinación entre dos sitios *frt*.

Se amplificó el cassette iTol2-Amp a partir del plásmido piTol2-Amp (Suster y cols 2011) usando los cebadores descritos pindigobac_itol2_fw y pindigobac_itol2_rev (Suster y cols 2011). Este cassette contiene las secuencias cis invertidas necesaria para la transgénesis con el sistema Tol2. Las bacterias SW105 con el BAC DK74B2 fueron inducidas a 42°C y electroporadas con el producto de PCR. Como control siempre electroporamos con la misma cantidad de producto de PCR bacterias donde la maquinaria de recombinación no ha sido inducida. Las bacterias, tanto inducidas como control, se crecen en medio con Cloranfenicol + Ampicilina. Se comprueba que no existen colonias en la placa control, o que el número de estas es muy inferior a la placa con las bacterias inducidas. AL menos 10 clones positivos se confirmaron por PCR y digestión (Suster y cols 2011). Obteniendo el BAC DK74B2-iTol2. La inserción del gen reportero GFP en el locus de *six2a* se llevó a cabo también mediante recombinación homóloga. Para este fin, se amplificó por PCR el cassette GFP-pA-FRT-kan-FRT desde el plásmido pBSKGFP-pA-FRT-kan-FRT (Suster ML y cols) usando los cebadores

Fwd_six2aHA1-Gfp:

TTAGATAGACATACAAGTACAAAGAGGGACGTTTATTTTTGAGACAAACCgccacc
atgGTGAGCAAGGGCGAGGAGCTGTTC y ReV_six2aHA2-frt-kan-frt: CAGA

CGCACGCCACTTGCTCTTGCGTAAAGCCGAATGTTGGAAGCATAGACCGCGTGT
AGGCTGGAGCTGCTTC. Así generamos un producto de PCR que contiene la secuencia codificante para la GFP seguida de una resistencia a kanamicina flanqueada por dos sitios idénticos *frt*, todo esto a su vez flanqueado por dos fragmentos de 50 pb homólogas a las regiones (-53, -3) y (+3,53) respecto del inicio de la traducción del gen *six2a*. Con este producto de PCR se electroporaron bacterias que contenían el BAC DK74B2-iTol2. Las bacterias se crecieron en medio con Ampicilina y Kanamicina. Se analizan al menos 10 colonias de la placa inducida para comprobar el evento de recombinación. Se realizan PCRs sobre los dos brazos de recombinación, usando un cebador interno a nuestro constructo y uno externo para cada brazo (Checksix2aHA1-Fwd:

TGTGCCTCTCTTCACCCGGTG,

Checksix2aHA1-ReV:

CAGCTCCTCGCCCTTGCTCAC,

Checksix2aHA2-Fwd: GAAGCAGCTCCAGCCTACACG,

Checksix2aHA2-ReV: GGGAGAACTGGTGGCTTTTCGAG)

Los clones positivos son además comprobados mediante digestiones con enzimas de restricción que produzcan bandas diferenciales con respecto a los clones donde no se ha producido la recombinación. Tras este paso, se incluye un paso de escisión del cassette de Kanamicina mediante inducción del gen *flpe*. Para esto se cultivan la bacterias en un medio con L-arabinosa (Suster ML y cols 2011) y se plaquean en placas con Cloramfenicol + Ampicilina. Se seleccionan los clones sensibles a Kanamicina, haciendo réplicas de la placa

en medio con Cloramfenicol + Ampicilina + Kanamicina. Los clones sensibles fueron además confirmados por PCR usando cebadores específicos (flp_induction_Fwd: ACGAGCTGTACAAGTAAAGCGGC y flp_induction_ReV: CCGCGTGTAGGCTGGAGCTGC). El cassette mCheery_KanR (Bussmann J y cols 2011) se amplificó por PCR incluyendo 50 pb de homología con el sitio de inicio de la translación de *six3a* de forma análoga al caso de *six2a* (Fwd_six3aHA-mCherry: TCGTCGTTCTTTTTTCCTTCGCAAATTTCACTCTCTCTCAGGTCATTTCCACCATG GTGAGCAAGGGCGAGGAG, ReV_six3aHA_KanR: TTTGGCAGGAAGAAATGAGAGGGATAAAGCTCTAAAGGCGATCTGAAAACCTCA GAAGAACTCGTCAAGAAGGCG). Este producto de PCR se usó para introducir por recombinación homóloga el gen reportero mCherry bajo el control del promotor de *six3a* en el BAC DK74B2-*six2a*::GFP-*iTol2* y así generar el BAC DK74B2-*six2a*::GFP-*six3a*::mCherry-*iTol2*. Al igual que en los casos anteriores la inserción se comprobó mediante PCR y digestiones con enzimas de restricción (Checksix3aHA1-Fwd: GATTGGCAGGGCTGCCATGAC, Checksix3aHA1-ReV: CCTCGCCCTTGCTCACCATGG, Checksix3aHA2-Fwd: GCCTTCTTGACGAGTTCTTCTG, Checksix3aHA2-Rev: GAAGCGACCATAGGAAGCG).

4.2 Generación del BAC $\Delta 18\text{kb}$ DK74B2-*six2a*::GFP-*six3a*::mCherry-*iTol2* (DelBAC).

A continuación describimos el proceso para la obtención de una versión del BAC DK74B2-*six2a*::GFP-*six3a*::mCherry-*iTol2* que incluye la eliminación de una región de aproximadamente 18 Kb, situada entre los genes *six2a* y *six3a* (chr13:9,801,660-9,820,362). Dos fragmentos de 250 pb flanqueantes a la región de interés fueron amplificados usando los siguientes primers: Del_EcoRI_HA1-Fwd:

TTGAATTCGCGTTCTTAGCAAGCAGGAT and Del_PstI_HA1-ReV: TTCTGCAGCGTTTAGCATCCACTCGTAGC (fragmento HA1)

y Del_XhoI_HA2-ReV: TTCTCGAGCCCCAGCCAACCCTTATTTTCG (fragmento HA2). Estos productos de PCR fueron posteriormente clonados independientemente en el vector pCR8-GW-TOPO. El cassette de resistencia a espectinomicina (SpeR) fue amplificado usando los siguientes cebadores PstI_SpeR_Fwd: TTCTGCAGCTGAAGCCAGTTA and PstI_SpeR_ReV: TTCTGCAGTAGCTGTTTCCTG y clonado también en un vector pCR8-GW-TOPO. El fragmento HA1, el fragmento HA2 y el cassette SpeR fueron escindidos de sus vectores mediante digestión con las enzimas EcoRI + PstI, PstI + XhoI, y PstI respectivamente. Todos los fragmentos fueron purificados y ligados conjuntamente junto a un plásmido pCS2+ previamente cortado con las enzimas EcoRI y XhoI. Los clones pCS2+_HA1_SpeR_HA2 fueron seleccionados mediante digestión con las enzimas EcoRI y XhoI. Este clone se usó como plantilla de una PCR

usando los cebadores Del_EcoRI_HA1-Fwd and Del_XhoI_HA2-ReV. Este producto de PCR es el “cassette de eliminación”, y se usó para la recombinación con el BAC DK74B2-six2a::GFP-six3a::mCherry. La selección de los clones positivos portadores de la eliminación, $\Delta 18\text{kb}$ DK74B2-six2a::GFP-six3a::mCherry-iTol2, se hizo mediante el cultivo en placas de LB + Cloranfenicol + Espectinomicina. Las colonias resultantes fueron comprobadas por PCRs sobre los extremos de la inserción del cassette de eliminación, usando cebadores internos y externos (DelCheck_six2a_Fwd(Genom):

CCAACCCTTATTTTCAGGTCATGC, DelCheck_six2a_ReV (Spectin):

GTCCACTGGGTTCGTGCCTTC; DelCheck_six3a_Fwd (Spectin):

GGCGAGATCACCAAGGTAGTCG; DelCheck_six3a_ReV (Genom):

TGAAGGAGTACCGCGAGTGTT). Además se realizaron dos digestiones independientes con XhoI y NotI sobre las dos versiones del BAC para verificar la presencia de la eliminación.

4.3 Purificación y transgénesis

Para la transgénesis en pez cebra, se purificaron ambas versiones del BAC (full y $\Delta 18$) con el kit Nucleobond BAC 100 (Macherey-Nagel), resuspendiendo al final del proceso cada BAC en 100 μl de agua doblemente destilada libre de ARNasas. A continuación ambos BACs se fenolizaron siguiendo la instrucciones de Suster y cols 2011. La concentración final de cada BAC se ajustó a 250 ng/ μl . La mezcla ADN de BAC/ARN de Tol2, se realizó de la siguiente manera: 4 μl de BAC (250 ng/ μl), 4 μl de ARNm de Tol2 (250 ng/ μl), 10 μl KCl (0,4M) y 2 μl de una solución de Rojo Fenol (2,5 % p/v). En cada embrión se inyectaron entre 2 y 5 nl de esta mezcla directamente en el embrión en el estadio un una célula. 24 horas después de la inyección se seleccionaron aquellos embriones que presentaban expresión de los reporteros mCherry y GFP. Cuando estos embriones alcanzaron la madurez sexual se cruzaron con peces silvestres para la búsqueda de peces positivos para la transgénesis.

5. BÚSQUEDA DE SITIOS DE UNIÓN DE CTCF

La búsqueda de los sitios de unión a CTCF en los bordes topológicos calculados de los complejos *Six* y la determinación de su orientación se realizó de forma manual. Para ello nos servimos del buscador genómico de la Universidad de California- Santa Cruz (UCSC genome browser, <https://genome.ucsc.edu>, por su facilidad de uso, su rapidez y la gran cantidad de datos experimentales integrados que contiene) y los distintos conjuntos de datos genómicos públicos que encontramos para humano y ratón, o datos generados en nuestro laboratorio para pez cebra. Estos datos los complementamos con los proporcionados en el buscador genómico de la Universidad de Washington (WashU Epigenome Browser,

<http://epigenomegateway.wustl.edu>) donde podemos encontrar y visualizar los datos de HiC para de células madre embrionarias de humano y ratón.

En los complejos *Six* de humano y ratón disponemos directamente de datos de ChIP-seq de sitios de unión de CTCF en distintos tipos celulares y tejidos derivados del proyecto ENCODE (www.encodeproject.org) . Basándonos en los bordes topológicos que se desprenden de los datos públicos de HiC, buscamos sitios de unión a CTCF constitutivos, es decir, que estén presentes en todos o muchos tipos celulares o tejidos. Para calcular la orientación de dichos sitios de unión a CTCF, analizamos la secuencia correspondiente a los picos de ChIP-seq usando la herramienta de búsqueda en línea de la base de datos Jaspar (<http://jaspar.genereg.net>) que permite determinar en qué cadena se encuentra el elemento CTCF, utilizando para ello la matriz de peso. Para el caso de los complejos de pez cebra, buscamos en la región alrededor de los bordes topológicos calculados. En ausencia de datos de ChIP-seq, utilizamos datos de ATAC-seq generados en nuestro laboratorio (Gehrke y cols 2015) en embriones completos a 24 hpf. Buscamos mediante la herramienta Jaspar, y con un valor umbral superior a 10, sitios de unión a CTCF presentes en algún pico de ATAC-seq. Por último para el caso de erizo de mar nos basamos exclusivamente en las predicciones de sitios de unión de CTCF obtenidas con el uso de Jaspar en la región alrededor del borde topológico predicho en el único complejo *Six*.

-Estudio a escala global de la distribución de sitios de unión de CTCF a lo largo de los bordes topológicos presentes en humano y ratón

Para calcular la posición media de los sitios de unión a CTCF presentes en ambas cadenas (cada cadena representa una orientación) respecto de los bordes topológicos, buscamos los sitios de unión a CTCF localizados dentro de picos de ChIP-seq comunes en al menos cinco tejidos diferentes en ratón (Dunham y cols 2012) o constitutivos en humanos (Li y cols 2013), que se encuentran en una ventana de 100 kb alrededor de cada borde. Para ello usamos el programa Clover (Frith y cols 2004) con un valor umbral de 10. Luego representamos la distribución posicional a lo largo de esas 100 kb alrededor de los bordes de los motivos identificados en función de su orientación mediante el uso de diagramas de cajas horizontales. Para poder comparar directamente la densidad de motivos en una u otra orientación en una posición concreta, dividimos las 100kb en ventanas de 5kb. Así podemos calcular el número de motivos de unión a CTCF en una determinada orientación en cada ventana, y compararlo con el número de motivos en la orientación opuesta en la misma ventana.

Como control hemos repetido el mismo análisis sobre 1000 ventanas de 100 kb centradas en promotores enriquecidos en sitios de unión a CTCF que no están involucrados en la formación de contactos de larga distancia. Este análisis se hizo tanto para humano como para ratón.

6. HIBRIDACIÓN IN SITU

6.1 *Pez cebra*

La detección de ARN en embriones completo del pez cebra, se realizó siguiendo protocolos anteriormente descritos (Tena y cols 2007, Jowett y Lettice 1994).

Los vectores con los ADNc de los genes *six2a*, *six3a*, *six2b* y *six3b* fueron suministrados por el laboratorio de la Dra. Paola Bovolenta (Centro de Biología Molecular Severo Ochoa). El ADNc del resto de genes *Six* de pez cebra estudiados en este trabajo fueron amplificados a partir del ADNc total obtenido por retrotranscripción de extracciones de ARNm de embriones completos de pez cebra a 36 hpf, siguiendo protocolos estándar. Después fueron clonados en vectores pGEM-T y secuenciados. Los cebadores usados para amplificar estos ADNc pueden consultarse en el ANEXO 5.

La sonda de ARN necesaria en cada caso se obtuvo mediante la transcripción del ADN complementario (ADNc) correspondiente. El vector con el ADNc fue linearizado con una enzima de restricción con sitio diana en el extremo ATG e, idealmente, sin ningún otro. Tras la purificación del ADN linearizado con fenol y cloroformo y su posterior precipitación, se resuspendió en agua mQ DEPC y se procedió a la transcripción. Para ello, se utilizaron 40 U de la polimerasa de ARN T7 o SP6 (nº cat. 10881767001 y 10810274001, Roche Applied Science) en un volumen final de 25 µl de búfer de transcripción, junto con 10 mM DTT (Ditiotreitol), 0,25 mM NTPmix con digoxigenina (*Digoxigenin-11-UTP*, nº cat. 11209256910, Roche Applied Science) como molécula marcadora y 10 U RNase Inhibitor (nº cat. 03 335 399 001, Roche). Tras la eliminación de los restos de ADN del vector con DesoxirribonucleasaI (DNaseI, nº cat. 85444320, Boehringer Mannheim), el ARN transcrito fue purificado mediante columnas de dextrano (*Sephadex G-50 medium*, nº cat. G-50150, Sigma-Aldrich) y cuantificado mediante un espectrofotómetro. Los embriones fueron rehidratados mediante lavados en soluciones con concentración 40 creciente de 0,1% PBTw, tratados con 10 µg de Proteinasa K, e hibridados, tras un paso previo de prehibridación, con la sonda a 2 ng/µl. Los restos no hibridados fueron lavados con soluciones de concentración decreciente de formamida y, posteriormente, los embriones fueron bloqueados (pez cebra: PBT, 2% suero de cabra (nº cat. B15- 035, PAA laboratories, Inc), 2 mg/ml albúmina de suero bovino (nº cat. A3294, Sigma-Aldrich)).

La detección de la molécula marcadora se consiguió mediante la incubación con anticuerpos anti-digoxigenina conjugados con la enzima fosfatasa alcalina (nº cat. 11093274910, Roche Applied Science) a concentración 1/2000. La señal fue revelada con NBT-BCIP (nº cat. 11681451001, Roche Applied Science) en búfer alcalino (100mM Tris-HCl pH 9.5, 50mM MgCl₂, 100mM NaCl, 0.1% Tween-20). Cuando la señal alcanzó la intensidad adecuada, la reacción fue detenida mediante lavados en PBS. Finalmente, los embriones fueron fijados y conservados en PFA al 4% a temperatura ambiente.

6.2 Ratón

Las sondas de los genes *Six2* y *Six3* fueron proporcionadas por el Dr. Pascal Maire (Institute of Biomedical Research, Paris) y el Dr. Tristán Rodríguez (Facultad de Medicina, Imperial College London, Londres) respectivamente. La sonda para el gen *Six6* ha sido descrita anteriormente (Gallardo y cols 1999). Las sondas para los genes *Six1* y *Six4* fueron generadas por amplificación reversa de extractos de RNA de embriones completos (estadio 9.5 dpc) usando las siguientes parejas de cebadores: *Six1*-F: CTCGGTCCTTCTGCTCCA, *Six1*-R: TCGCTGCTACCCTAACCG, *Six4*-F: CCTTTATGCTGTTAGTCCAGGG, *Six4*-R:

CCGCCTTATAGAACGTTTGTGAC, tal y como se describe en el “Allen Brain Atlas” (www.brain-map.org). Los fragmentos amplificados fueron aislados y clonados en el vector pSpark (Canvas Biotech). Las ribosondas fueron preparadas usando la ARN-polimerasa T7 sobre los clones linealizados siguiendo las instrucciones del productor (Roche Diagnostics). La hibridación in situ se realizó sobre embriones completos de ratón de la estirpe Cba X C57/(B6) usando ribosondas marcadas con digoxigenina como se ha descrito previamente (Summerbell y cols 2000.). Para el proceso de hibridación in situ se usó el robot InsituPro (Intavis).

6.3 Erizo de mar

La hibridación in situ sobre embriones completos de *Stroglyocentrotus purpuratus* ha sido descrita con anterioridad (Ransick y cols 1993). Las sondas para los genes *Sp-Six1/2* y *Sp-Six3/6* fueron suministradas por el laboratorio del doctor Eric H. Davidson (California Institute of Technology). Brevemente el proceso es como sigue. Los embriones de erizo de mar son fijados en una solución de glutaraldehído. Los embriones fijados son incubados en búfer de hibridación (50% (vol/vol), formamida, 5x SSC, 1x Denhardt's, 1mg/ml de ARNt de levadura, 50ng/ml de heparina, y 0,1% de Tween-20) con 0,5 ng/μl de ribsonda marcada con digoxigenina a 60°C durante 18h. A continuación se realizaron una serie de lavados: búfer de hibridación, 2x SSCT (2x SSC y 0,1% Tween-20), 0,2x SSCT y 0,1x de SSCT. Cada lavado durante 20 minutos a 60°C. A continuación, se realiza una incubación a temperatura ambiente con una dilución 1:1000 del anticuerpo anti-digoxigenina conjugados con fosfatasa alcalina (Roche). Tras la incubación los embriones son intensamente lavados: 6 lavados con búfer MABT (ácido málico 0.1M, NaCl 0,15M y 0,1% de Tween-20) y dos lavados con búfer AP (Tris-HCl 100mM pH 9,5, NaCl 100mM, MgCl₂ 50mM y levamisole 1mM). Para la reacción de tinción se usó BCIP y NBT.

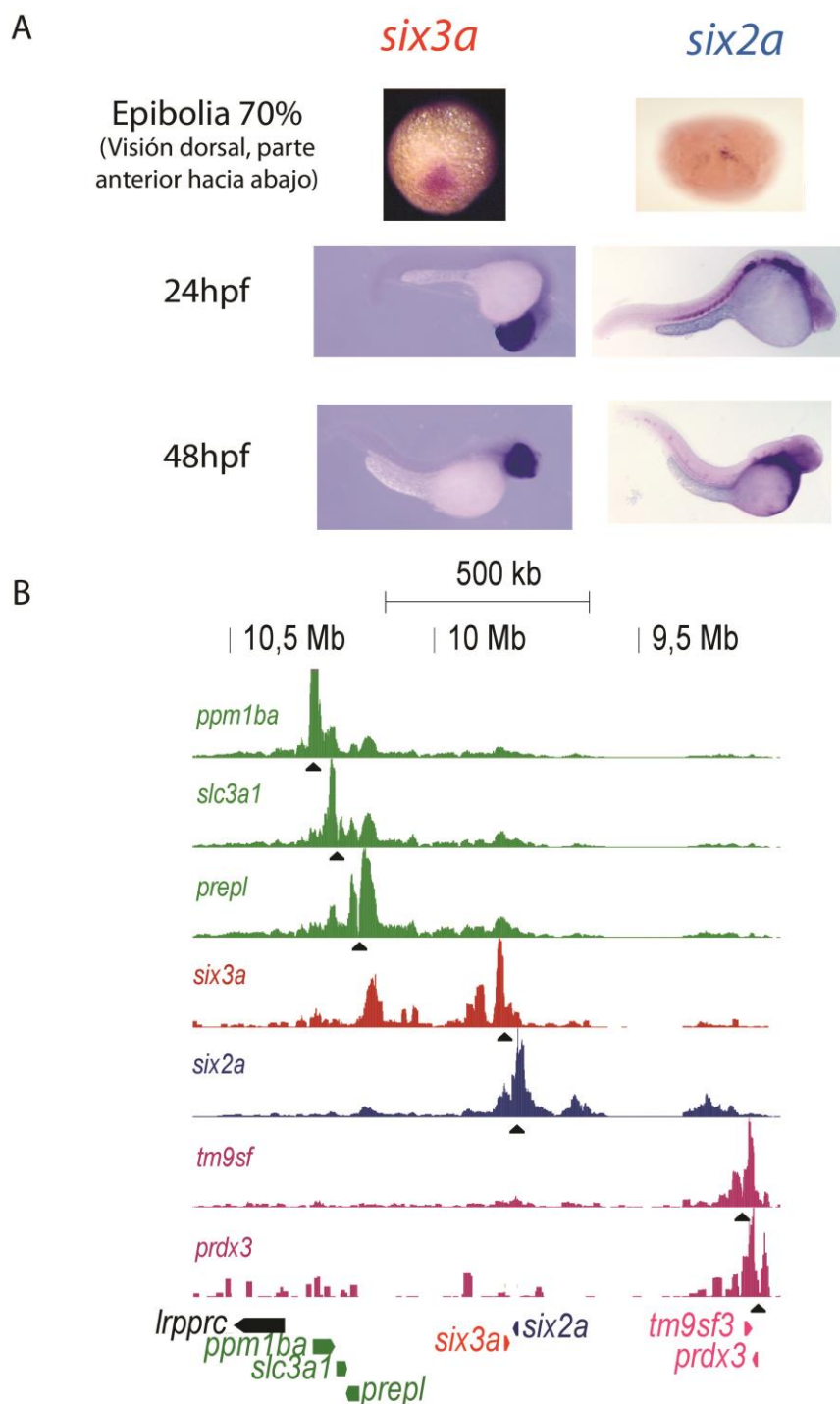
Resultados

Resultados

1. CARACTERIZACIÓN DEL ENTORNO REGULADOR DE LOS GENES *six2a* Y *six3a* EN EMBRIONES DE PEZ ZEBRA A 24 HPF.

Los genes *six2a* y *six3a*, son esenciales para el correcto desarrollo de multitud de estructuras durante el desarrollo embrionario de pez cebra, entre las cuales destacan el cerebro anterior, la retina, el mesodermo facial, el músculo esquelético o los riñones. Estos genes se expresan de forma diferente durante el desarrollo, produciéndose poco solapamiento en sus dominios de expresión (Fig. RE_1 A). La hipótesis de partida fue que esta transcripción diferencial se podría deber a que ambos genes se encuentran bajo paisajes reguladores distintos, a pesar de su cercanía física en la secuencia de ADN. Por tanto para testar esta hipótesis decidimos aplicar la técnica de 4C (Circular Chromatine Conformation Capture) a los promotores de ambos genes, lo que nos permitiría obtener sus perfiles genómicos de contacto. De esta forma podemos comprobar si, en efecto, cada promotor interactúa con regiones independientes de la cromatina. Para ello usamos embriones completos de pez cebra a 24hpf, estadio en que ambos genes se encuentran claramente activos. Para aportar una mayor información estructural del locus y poder relativizar los resultados, decidimos incluir 4Cs sobre los promotores de algunos genes circundantes, incluyendo: *ppm1ba*, *slc3a1* y *prepl* aguas arriba de *six3a*; y *tm9sf3*, y *prdx3* aguas abajo de *six2a*. Los patrones de contacto proporcionados por los 4C revelan que *six3a* tiene una tendencia a interactuar con la región aguas arriba del promotor extendiéndose sus contactos más de 500kb y llegando a las inmediaciones del gen *lrpprc* (Figura RE_1B). Aguas abajo de su promotor, los contactos de *six3a* llegan hasta justo antes del promotor del gen *prdx3* (extendiéndose otras 500kb aproximadamente), sin embargo, la frecuencia de contactos en esta región es cualitativamente mucho menor que aguas arriba del promotor. Por otro lado, los contactos de *six2a* se extienden aproximadamente a lo largo de la misma región cromosómica que *six3a* pero con una distribución en la frecuencia de interacción diametralmente opuesta. En este caso la mayor parte de los contactos de *six2a* se producen aguas abajo del promotor (en dirección al gen *prdx3*).

Los genes *ppm1ba*, *slc3a1* y *prepl* se encuentran en el entorno de *six3a*, esta posición u orden relativo (sintenia) respecto de *six3a*, también aparece en otros vertebrados como



11 Fig RE_1

Fig. RE_1: La expresión diferencial de los genes *six3a* y *six2a* correlaciona con la división del locus en dos dominios físicos separados. A) Expresión de los genes *six3a* y *six2a* en distintos momentos del desarrollo. B) Perfiles de 4C-seq obtenidos a partir de embriones completos a 24hpf. Se muestran perfiles de contacto suavizados de los genes *six3a* (rojo) y *six2a* (azul) y de los genes circundantes: *ppm1ba*, *slc3a1* y *prepl* en verde y *tm9sf* y *prdx3* en rosa. La posición del punto fijo usado en cada 4C-seq se muestra mediante triángulos negros debajo de cada 4C. La última fila representa los genes en su contexto genómico.

el ratón (*Ppm1b*, *Slc3a* y *Prepl* aparecen junto a *Six3*). Esto sugiere que dichos genes han estado ligados al locus *six3/six2* probablemente en gran parte del linaje de vertebrados. Este hecho concuerda con su integración dentro del paisaje regulador de *six3a* dado por la señal de 4C desde su promotor. Por otro lado, el perfil de contactos de estos genes se expande con una señal basal pero continua hasta las inmediaciones de los genes *six3a/six2a*. Estos datos muestran como la relación sinténica de *ppm1ba*, *slc3a1* y *prepl* con *six3a* se acompaña de interacciones físicas, sugiriendo una relación reguladora en cis. En contraposición, tenemos los genes *prdx3* y *tm9sf3* en el extremo opuesto del locus. La relación posicional de estos genes con *six2a* es específica del pez cebra (estos genes no se encuentran asociados a los genes *Six* ni en humanos ni en ratón), por lo que no se puede hablar de sintenia. El perfil de contactos de *six2a* no muestra ninguna señal clara que abarque a *prdx3* o *tm9sf3*. Por otro lado, los perfiles de contactos de *prdx3* y *tm9sf3* no se extienden de manera continua dentro del paisaje regulador *six2a* (a excepción de picos aislados del 4C de *tm9sf3*, señal producida probablemente por una baja calidad de la librería). Esto indica que la falta de relación sinténica entre los genes se correlaciona con la ausencia o muy baja presencia de interacciones físicas entre ellos y el paisaje regulador de *six2a*. En consecuencia, estos datos sugerirían que no existe, o sería muy débil, una relación reguladora en cis entre el locus *six3/six2* y los genes *prdx3* y *tm9sf3*.

Los patrones de contacto de *six3a* y *six2a* en embriones de pez cebra a 24hpf apuntan a que el locus estudiado podría estar dividido en dos dominios topológicos diferenciados. Uno marcado por un dominio en la frecuencia de contactos del promotor del gen *six3a* y el otro marcado por un dominio en la frecuencia de contactos del promotor del gen *six2a*. Si esta suposición es cierta, existiría un punto que marcaría el límite entre ambos dominios topológicamente asociados (TADs), que se correspondería con un borde topológico. Ese punto concreto, se caracterizaría por:

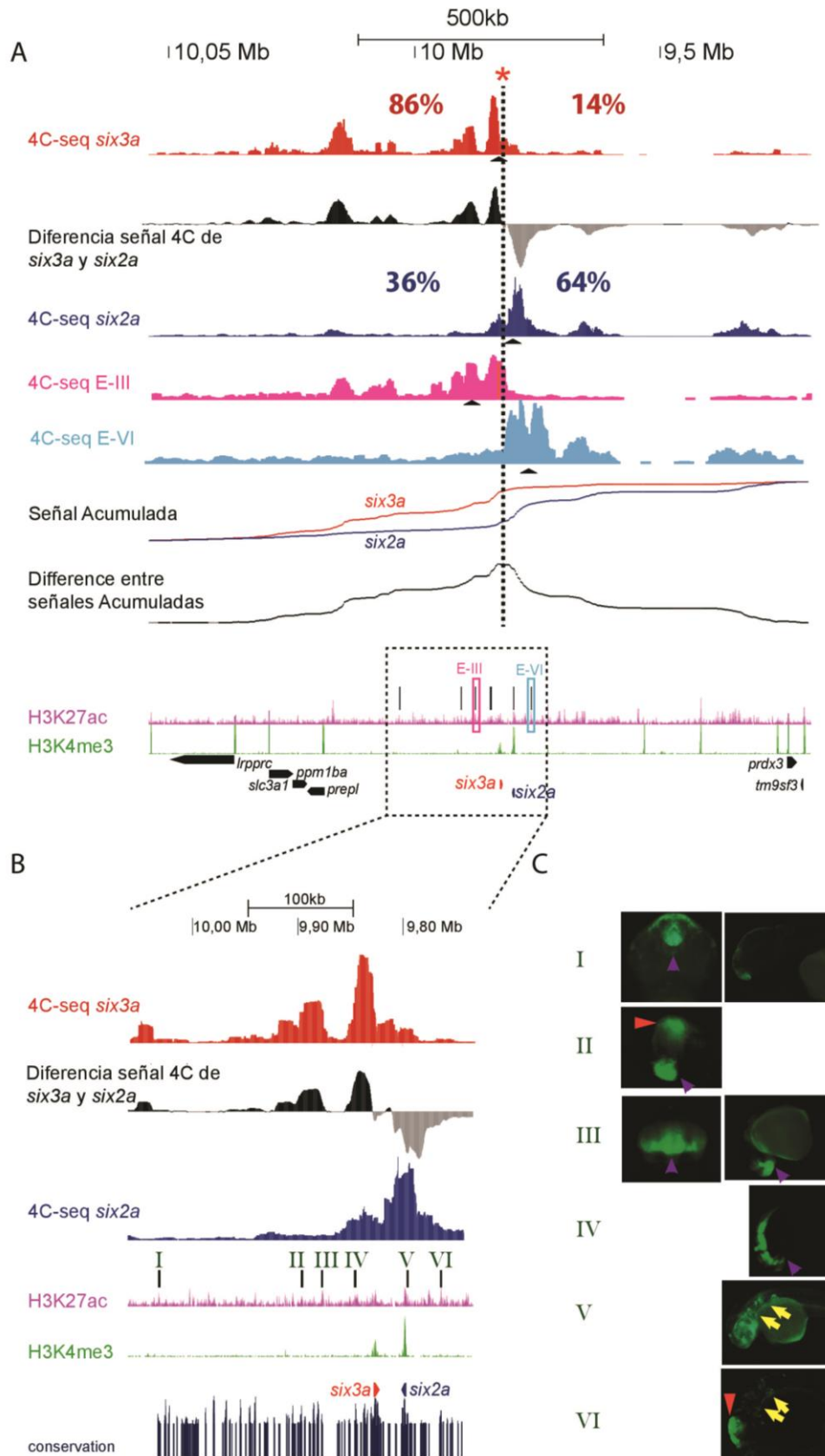
- 1) Ser el punto en el cuál, la diferencia en la distribución de contactos entre el promotor de *six3a* y el promotor de *six2a* es máxima a un lado y a otro del mismo. Esto es, que aguas arriba de dicho punto la diferencia entre el porcentaje de contactos del promotor de *six3a* y de *six2a* sea máxima, y que ocurra lo mismo pero a la inversa aguas abajo de dicho punto.
- 2) Ser el punto a partir del cual, a un lado la frecuencia relativa de contacto de *six3a* en cada fragmento DpnII-DpnII tiende a ser mayor que la frecuencia relativa de contacto de *six2a*, mientras que al otro ocurre al contrario.

Para inferir cual sería dicho borde topológico debemos encontrar una posición que cumpla ambas premisas simultáneamente. Para encontrar el punto que cumple la primera premisa, decidimos calcular el porcentaje acumulado de contactos o “lecturas” (“reads”) a lo largo del locus *six3/six2* (si bien la delimitación exacta del locus ha sido manual, la extensión del mismo abarca la gran mayoría de los contactos significativos de los promotores de *six2a* y *six3a*) (Fig. RE_2_A, sexto panel). El punto que buscamos será aquél donde la diferencia entre el porcentaje acumulado de “lecturas” de ambos promotores sea máxima (Fig. RE_2_A, último panel). Para encontrar el punto que cumpla la segunda premisa, bastará con normalizar los datos de cada 4C (ver sección Materiales y Métodos), y calcular la diferencia entre el patrón de contacto de *six3a* y el de *six2a* (Fig. RE_2_A, segundo panel). Como podemos observar, hay un punto, aguas arriba del cuál la frecuencia de contacto de *six3a* es mayor que la de *six2a* para la mayoría de los fragmentos DpnII-DpnII, y aguas abajo del cuál ocurre lo contrario. Vemos como el punto que cumple la primera premisa así como el que cumple la segunda, son el mismo (Fig. RE_2_A, asterisco y línea discontinua). Este punto representaría el “borde topológico inferido o predicho” (BTP) y marca el punto de inflexión donde la frecuencia de contactos de *six3a* decae abruptamente a la vez que la frecuencia de contactos de *six2a* se incrementa de manera igualmente abrupta, a pesar de la gran cercanía de ambos promotores (~30kb) en comparación con la extensión del locus (1 Mb).

Si tomamos este “borde topológico predicho” o BTP como punto de referencia, observamos como aguas arriba de dicho borde *six3a* tiene en torno al 86% de sus contactos, mientras que *six2a* tan sólo tiene el 36%. La distribución de los contactos aguas abajo del BTP es opuesta, siendo del 14% del total de contactos para *six3a* y del 64% para *six2a* (Fig. RE_2_A). Esto sugiere una partición física de locus que contiene este complejo de genes (*six2a-six3a*), en dos dominios topológicos distintos.

Para comprobar si esta partición física del locus en dos dominios físicos, se corresponde con una partición reguladora, decidimos realizar un estudio de elementos potenciadores a ambos lados del BTP. La finalidad es comprobar si el dominio físico correspondiente a *six3a*, contiene elementos con actividad potenciadora tipo *six3a*, y si ocurre lo mismo con el dominio correspondiente a *six2a*.

Resultados



12 Fig RE_2

Fig. RE_2: Un borde topológico separa el locus *six2a-six3a* en dos paisajes reguladores distintos.

A) Cálculo del Borde Topológico Predicho (BTP) en el locus *six3a-six2a*. El primer panel, el tercero el cuarto y el quinto muestran 4C-seq hechos sobre embriones de pez cebra completos a 24hpf desde el promotor de *six3a*, de *six2a*, desde potenciador E-III y desde el potenciador E-VI respectivamente. El sexto panel muestra la señal acumulada de los 4C-seq de *six2a* y *six3a*, la diferencia entre estas señales acumuladas se muestra en el último panel. El asterisco marca el punto donde esta diferencia es máxima. Este punto coincide con el punto en el que se produce un cambio en la dominancia relativa de contactos de *six3a* respecto a *six2a*, evento mostrado en el segundo panel como diferencia entre las señales de 4C-seq de *six3a* y *six2a*. Este punto es el Borde Topológico Predicho (BTP), el porcentaje de señal a cada lado del BTP de los 4C-seq de *six3a* y *six2a* aparece en la parte superior de cada 4C. En la parte inferior aparecen los perfiles de ChIP-seq para las modificaciones de histonas H3K27ac y H3K4me3 en embriones completos de pez cebra a 24hpf. Las líneas negras corresponden a elementos potenciadores caracterizados incluyendo E-III y E-VI **B)** Vista ampliada de la región contenida en el BAC DKEY-74B2 que se usará posteriormente. Vienen detallada la posición de los elementos I a VI. **C)** Embriones de pez cebra a 24hpf transgénicos que muestran la expresión de GFP dirigida por los elementos de ADN en cis, testados para comprobar su actividad potenciadora. Los elementos II y VI se probaron usando el vector Jlz48 (la expresión control en el cerebro medio está marcada por una flecha roja), para el resto de elementos se usó el vector ZED. Los elementos I a IV dirigen la expresión de GFP en territorios compartidos con *six3a* (cerebro anterior y retina ventral, flechas moradas), mientras que los elementos V y VI lo hacen en territorios compartidos con la expresión de *six2a* (arcos branquiales y aleta pectoral, flechas amarillas).

La selección de elementos potenciadores candidatos se llevó a cabo basándonos en los perfiles de señal histona 3 lisina 27 acetilada (H3K27ac, marca asociada con elementos cis-reguladores activos) generadas mediante ChIP-seq sobre embriones completos de pez cebra a 24 hpf (Bogdanovic y cols 2012). Se priorizaron aquellas regiones que además estuviesen conservadas en vertebrados (basándonos en los datos de conservación del buscador genómico de la Universidad de California – Santa Cruz), y que estuvieran incluidas en la región abarcada por el BAC (cromosoma artificial de bacteria del inglés “Bacterial Artificial Chromosome”) DK74B2 que se usará posteriormente. Para testar la actividad potenciadora, los elementos candidatos se clonaron bien en el vector ZED o en el vector Jlz48 (ver sección Materiales y Métodos), y se realizó transgénesis en pez cebra con el sistema tol2. De los 9 candidatos seleccionados, 6 incluidos en la región del BAC DK74B2 mostraron actividad potenciadora relacionada con los genes *six3a* o *six2a* a 24 hpf (Fig. RE_2_B). Cuatro elementos mostraron actividad potenciadora reproducible en embriones transgénicos estables (Elementos I, III, IV y V) y dos lo hicieron en individuos transgénicos transitorios (Elementos IV y VI), el análisis del resto de elementos fue no concluyente (Elementos N1, N2 y N3). Los cuatro elementos situados aguas arriba del BTP (elementos I, II, III y IV) muestran actividad transcripcional en territorios donde *six3a* es expresado, tales como el cerebro anterior o la retina ventral (Fig. RE_2_C). Por otro lado los elementos aguas abajo del BTP (elementos V y VI) muestran actividad transcripcional en territorios que

correlacionan con la expresión endógena de *six2a*, como es el caso de la vesícula ótica y el mesénquima branquial. Esto sugiere que la partición física del locus en 2 dominios topológicos subyace a una partición de locus en 2 paisajes reguladores distintos.

La partición del locus *six3a-six2a* en dos dominios topológicos ayudaría al direccionamiento de los elementos potenciadores hacia sus promotores diana y a evitar contactos promiscuos entre elementos potenciadores de un dominio topológico y el promotor del gen situado en el otro dominio. Si esta hipótesis es cierta los elementos potenciadores aguas arriba del BTP, contactarían con mayor frecuencia el promotor de *six3a* que con el promotor de *six2a*, ocurriendo a la inversa para los elementos potenciadores aguas abajo del BTP. Para comprobar esto, realizamos 4Cs usando como puntos fijos los elementos III (aguas arriba del BTP) y VI (aguas abajo del BTP) que tienen actividad reguladora similar a la expresión endógena de *six3a* y *six2a* respectivamente (Figura RE_2_C). El perfil de contacto del elemento III reveló una fuerte interacción con la región que abarca al promotor de *six3a* (Fig. RE_2_A, cuarto panel). Podemos observar como los contactos del elemento III más allá del BTP son casi inexistentes, mientras que hacia la dirección opuesta lo contactos alcanzan varios cientos de kilobases. Además el elemento III comparte en gran medida el perfil de contacto de *six3a*, sugiriendo que existen multitud de interacciones entre elementos potenciadores del entorno regulador de *six3a*. Un comportamiento análogo encontramos cuando analizamos el perfil de contactos del elemento VI (Fig. RE_2_A, quinto panel). Este elemento contacta fuertemente con el promotor de *six2a*, con el que también comparte otros muchos contactos. Además los contactos del elemento VI con regiones al otro lado del BTP son también de una frecuencia mucho menor. Estos resultados confirmarían el borde topológico que propusimos anteriormente, y que la organización del complejo *six3a-six2a* en dos TADs diferentes, efectivamente, soporta la partición del locus en dos paisajes reguladores fuertemente aislados.

2. ENTORNO REGULADOR DE LOS GENES *six3a* Y *six2a* DURANTE EL DESARROLLO DEL PEZ ZEBRA.

Hemos podido comprobar como la utilización de la técnica de 4C en embriones completos revela que a 24 hpf el locus del complejo *six3a-six2a* se encuentra dividido en dos dominios físicos que aíslan el entorno regulador de cada gen. En este estadio de desarrollo ambos genes se encuentran activos en diversos territorios del embrión (Fig. RE_1_A). Para ver si la

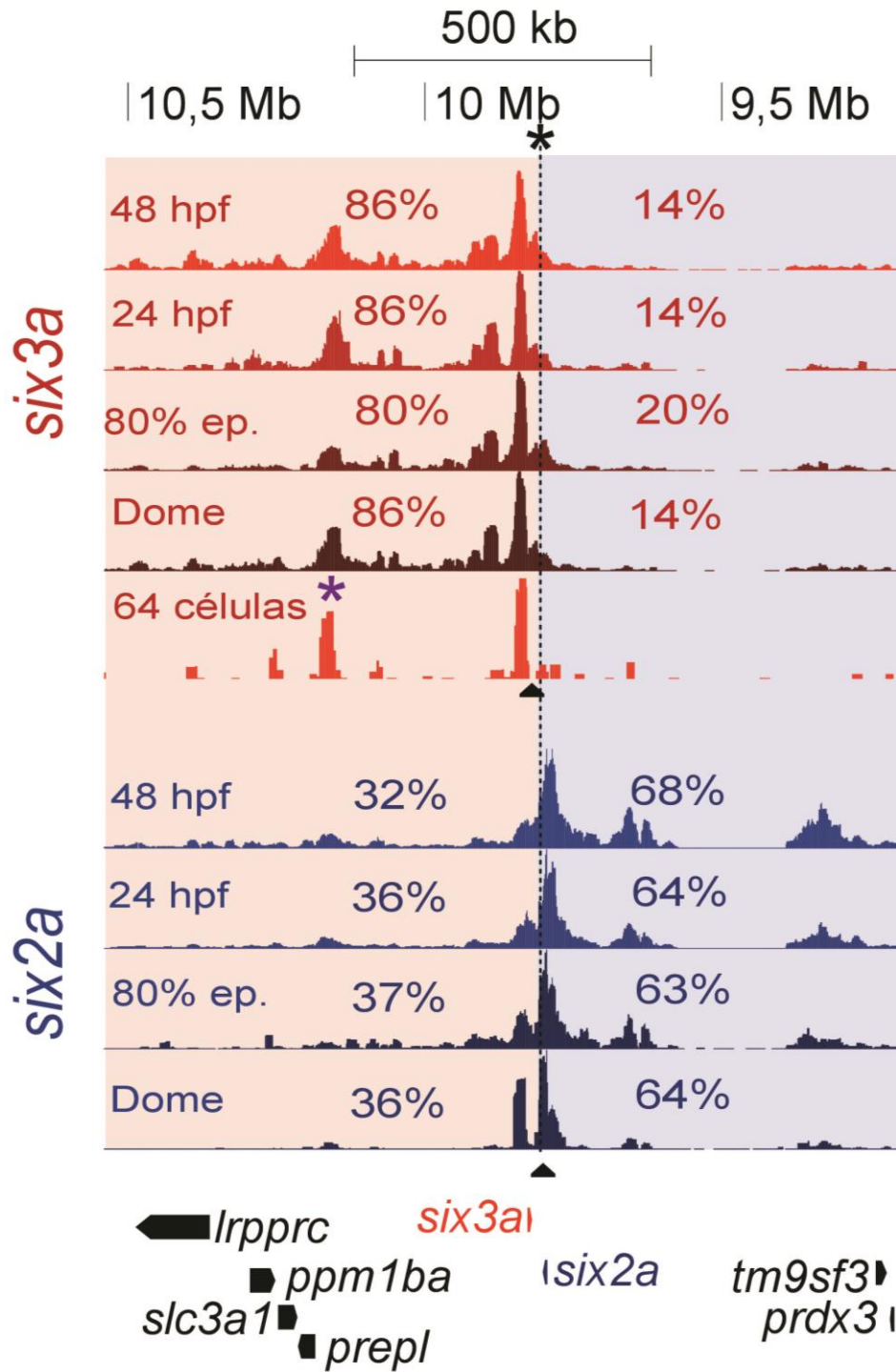
organización tridimensional que observamos en el locus es dependiente del estado regulador de los genes, decidimos realizar 4Cs sobre embriones completos desde los promotores de *six3a* y *six2a* en distintos momentos del desarrollo embrionario del pez cebra (Fig RE_3). Empezamos desde el estadio de blástula (“Dome” o “cúpula” en español), donde ninguno de los genes está activo, pasando por 80% de epibolia (donde ambos genes han comenzado ya a transcribirse) y 24 hpf, terminando en 48 hpf (“pharyngula” o faringula en español), donde la mayoría de los órganos del embrión están formados y ambos genes continúan expresándose (Fig. RE_1_A).

La inspección visual de los perfiles de contactos de los promotores de *six3a* y *six2a* en los diversos momentos del desarrollo no muestran ningún cambio evidente (Fig RE_3). De hecho ambos perfiles se mantienen prácticamente constantes a lo largo del desarrollo del pez zebra, incluso en el momento de blástula donde ninguno de los genes ha sido activado aún. Para cuantificar el grado de similitud entre los perfiles de contacto en los distintos momentos de desarrollo, calculamos el coeficiente de correlación de pearson (p) entre el perfil de contacto de un gen en distintos momentos del desarrollo, y lo comparamos con el coeficiente de correlación que existe entre réplicas biológicas. Dada la complejidad y el coste experimental de los 4Cs y la similitud visual entre los distintos estadios de desarrollo, sólo realizamos réplicas biológicas de los 4Cs obtenidos a 24hpf (para pez zebra, en el caso de ratón que veremos más adelante, las réplicas se realizaron para el estadio de 14.5 dpf y para erizo de mar a las 48 hpf). Por otro lado, como la reproducibilidad de los 4Cs a veces se ve muy influenciada por la calidad de la muestra o la idoneidad de los cebadores, en vez de comparar los coeficientes de correlación individualmente, comparamos la media y distribución de los coeficientes de correlación obtenidos para todas las réplicas biológicas (de pez zebra, ratón y erizo de mar) con la media de los coeficientes obtenidos cuando comparamos distintos momentos del desarrollo (todas las comparaciones posibles) (Fig RE_9 A y C). Basándonos en estos coeficientes de correlación, cuando vemos dos perfiles de 4C de un gen no podemos distinguir si se trata del patrón de contactos de dos réplicas biológicas o de un mismo gen en distintos momentos del desarrollo. En consecuencia cuando inferimos el BTP (“borde topológico predicho”) entre *six3a* y *six2a*, en los distintos momentos del desarrollo siempre obtenemos la misma posición (Fig RE_3, asterisco y línea discontinúa), y cuando calculamos el porcentaje de señal a un lado y a otro del BTP para cada gen siempre obtenemos la misma distribución: 85% de la señal aguas arriba del BTP y 15% aguas abajo para *six3a*, y el 35% aguas arriba y el 65% aguas abajo para *six2a*. Estos datos, revelan que la

bipartición del complejo *six2a-six3a* podría ocurrir durante todo el desarrollo embrionario, independientemente de la actividad transcripcional de estos genes.

Para tratar de ver hasta dónde llega la conservación del perfil de contactos de estos genes en el desarrollo embrionario, decidimos hacer 4Cs sobre embriones con entre 8 y 64 células, mucho antes de cualquier activación transcripcional zigótica. Sin embargo esta aproximación experimental tiene una limitación: obtener las 10 millones de células que necesita una muestra estándar para 4C. En estos estadios de desarrollo se necesitarían entre 1250000 y 150000 embriones, unas cantidades difíciles de obtener. Así que, como prueba de concepto, partimos de unos 5000 embriones, entre 250 y 30 veces menos de la cantidad de material estándar. Como consecuencia de esta baja cantidad de cromatina fijada (y por ende, de contactos), no obtuvimos producto de PCR para el caso de *six2a*, y para *six3a* obtuvimos un perfil de contacto discontinuo (Fig. RE_3). Sin embargo, podemos observar como uno de los principales contactos distales de *six3a* en los límites de su entorno regulador, está ya presente aún en ausencia de la activación de la transcripción en el cigoto (Fig. RE_3). Si bien este dato requiere de una confirmación experimental, esto indicaría que el dominio topológico que sostiene el entorno regulador de *six3a*, esta ensamblado desde el inicio de la embriogénesis.

Dado que los contactos estructurales pueden estar enmascarando los contactos funcionales, hicimos un intento para averiguar si podría existir una configuración espacial distinta en función del estado regulatorio de los genes *six3a* y *six2a*. Para ello llevamos a cabo una nueva aproximación, a la que hemos llamado ChIP-4C (ver sección Materiales y Métodos). El principio básico es tratar de enriquecer la muestra en cromatina asociada a un estado regulatorio específico. Existen modificaciones postraduccionales de histonas que correlacionan con un estado regulador particular de los elementos de ADN asociados a ellas. Así tenemos la modificación H3K27ac que correlaciona con elementos cis-reguladores activos, o la modificación H3K27me3 que correlaciona con promotores silenciados (Rada-Iglesias y cols 2011, Bogdanovic y cols 2012). Si durante el procesamiento de una muestra



13 Fig RE_3

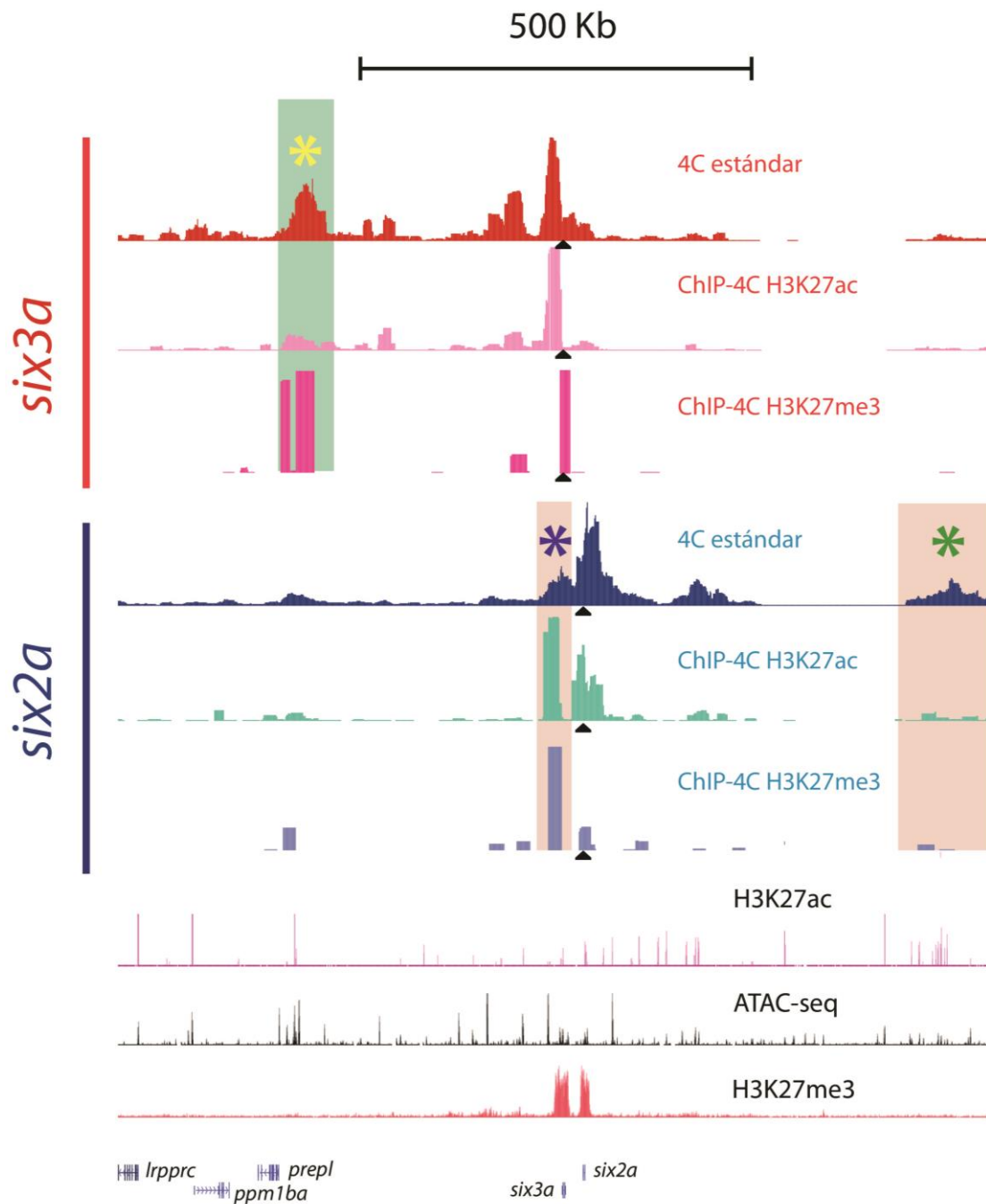
Fig. RE_3: Perfiles de 4C-seq de los genes *six2a* y *six3a* obtenidos en embriones completos en distintos momentos del desarrollo: Dome (o bóveda), epibolia al 80%, 24hpf y 48hpf. Además un perfil extra en estadio de 64 células para *six3a*. El asterisco morado muestra el pico distal más significativo que se encuentra presente al inicio de la embriogénesis. El asterisco negro y la línea discontinua muestran la localización del Borde Topológico Predicho. Los porcentajes se corresponden al porcentaje de señal de cada 4C-seq a un lado y al otro del BTP.

de 4C, incluimos un paso de inmunoprecipitación de cromatina usando anticuerpos contra estas modificaciones de histonas, esperamos obtener una muestra final enriquecida en cromatina asociada a estas marcas epigenéticas. En consecuencia tendríamos un enriquecimiento en interacciones asociadas a elementos reguladores activos (H3K27ac) o bien en interacciones asociadas a promotores silenciados (H3K27me3).

Obtuvimos 2 muestras de ChIP-4C usando embriones completos a 24hpf, una usando un anticuerpo contra H3K27ac y otra contra H3K27me3. Para cada muestra usamos 500 embriones, pero tras el proceso la concentración de cada muestra estuvo alrededor a 1ng/ μ l (unas 500 veces menos que una muestra de 4C normal) para el caso de la inmunoprecipitación con el anticuerpo anti H3K27ac, y alrededor de 0,2 ng/ μ l para el caso de la inmunoprecipitación de H3K27me3. Sobre cada muestra realizamos 4C desde el promotor de *six3a* y desde el de *six2a*. La falta de réplicas biológicas y de algún control del proceso de inmunoprecipitación, impidió cualquier tipo de interpretación sobre los resultados. Por otro lado la poca cantidad de muestra, requirió forzar la PCR especialmente en el caso de la muestra de H3K27me3, lo que dio lugar a un patrón de contactos muy discontinuo (Fig. RE_4).

El patrón de contactos de *six3a* en la muestra ChIP-4C-H3K27ac, es cualitativamente similar al patrón observado en el 4C original, si bien los contactos de *six3a* en el entorno de *six2a* (más allá del borde topológico) son menores. Por otro lado uno de los contactos distales de *six3a* más significativos en el 4C original, el cercano al gen *prepl* (Fig. RE_4), se encuentra reducido. En la muestra ChIP-4C-H3K27me3, sólo observamos dos contactos para el promotor *six3a*, probablemente debido a la poca cantidad de molde que pudimos usar para las PCRs. Uno de los contactos corresponde con el entorno inmediato de *six3a*, y el otro correspondería al contacto observado en el 4C original cercano al gen *prepl*.

En el caso de *six2a*, su patrón de contactos en la muestra ChIP-4C-H3K27ac destaca por una reducción en los contactos más distales (Fig. RE_4, asterisco verde). Hay que reseñar que en este contexto *six2a* parece tener un contacto muy fuerte más allá del borde topológico, en el entorno regulador más cercano a *six3a* (Fig. RE_4, asterisco morado). Este contacto más allá del borde topológico es la única señal de contacto que logramos captar para *six2a* en la muestra ChIP-4C-H3K27me3.



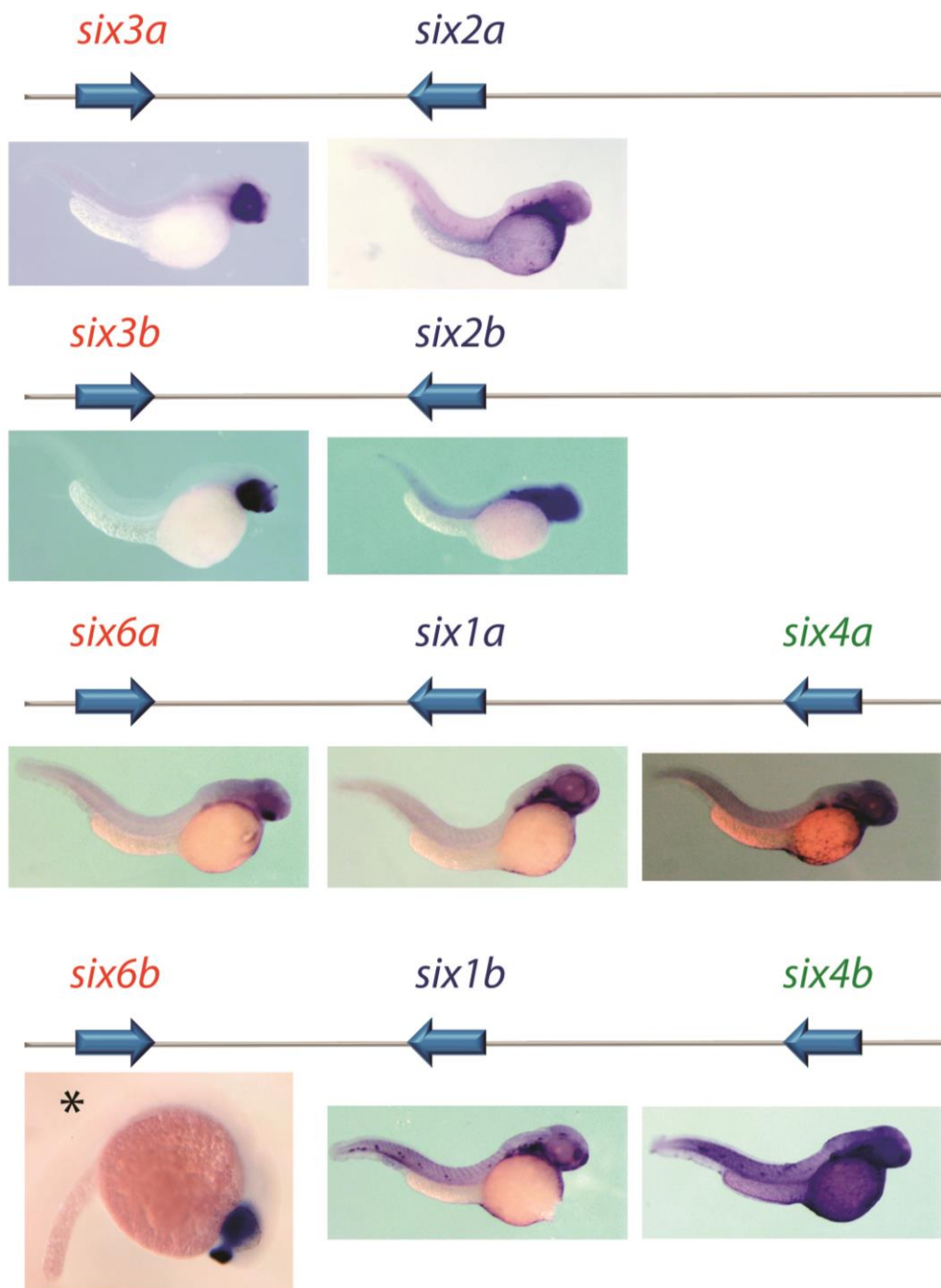
14 Fig RE_4

Fig. RE_4: Resultados de ChIP-4C para *six3a* y *six2a*. El panel superior muestra los perfiles para *six3a* a 24 hpf en embriones completos obtenidos de un 4C-seq estándar, un ChIP-4C de H3K27ac y un ChIP-4C H3K27me3. El siguiente panel muestra lo mismo para el gen *six2a*. Abajo se muestran los perfiles de ChIP-seq para H3K27ac y H3K27me3, así como el perfil de ATAC-seq, obtenidos en embriones completos de pez cebra a 24hpf. El asterisco amarillo muestra una región cuya señal de contacto disminuye en el ChIP-4C de *six3a*. El asterisco morado muestra una región que incrementa la señal de contacto en los ChIP-4C de *six2a* respecto al 4C-seq estándar, por otro lado el asterisco verde muestra una región con el comportamiento inverso.

El estudio de los perfiles de contactos de los genes *six2a* y *six3a* en distintos momentos del desarrollo del pez cebra, sugiere que el principio organizativo básico del locus permanece constante y es independiente del estado regulador de los genes. Si bien nuestra aproximación experimental usando embriones completos no nos permite descartar que se produzcan cambios puntuales en los perfiles de interacción que se escapen a la resolución de nuestro método. En cualquier caso, estos cambios puntuales no alterarían de forma significativa la estructura global del locus.

3. ORGANIZACIÓN TRIDIMENSIONAL DEL RESTO DE COMPLEJOS DE GENES *Six* DE PEZ ZEBRA

A continuación quisimos averiguar si el principio organizativo espacial del complejo *six3a-six2a* en dos TADs mantenidos a lo largo del desarrollo embrionario, estaba también presente en el resto de complejos de genes *six* de pez cebra. En el pez cebra encontramos otros 3 complejos de genes *six*: *six3b-six2b*, *six6a-six1a-six4a* y *six6b-six1b-six4b*. Como en el caso del complejo *six3a-six2a*, desde el punto de vista de los dominios de expresión, los genes *six* que componen un mismo complejo parecen estar divididos en dos entornos reguladores similares a los que encontramos en el complejo *six3a-six2a*. Sabemos que los genes *six3b*, *six6a* y *six6b*, tienen dominios de expresión similares a *six3a*, en este caso en el cerebro anterior o la retina ventral, mientras que los genes *six2b*, *six1a*, *six1b*, *six4a* y *six4b* tienen dominios de expresión similares entre sí y más parecidos a *six2a*, y encontramos expresión en la vesícula ótica, el mesénquima branquial o la línea lateral (Fig. RE_5).

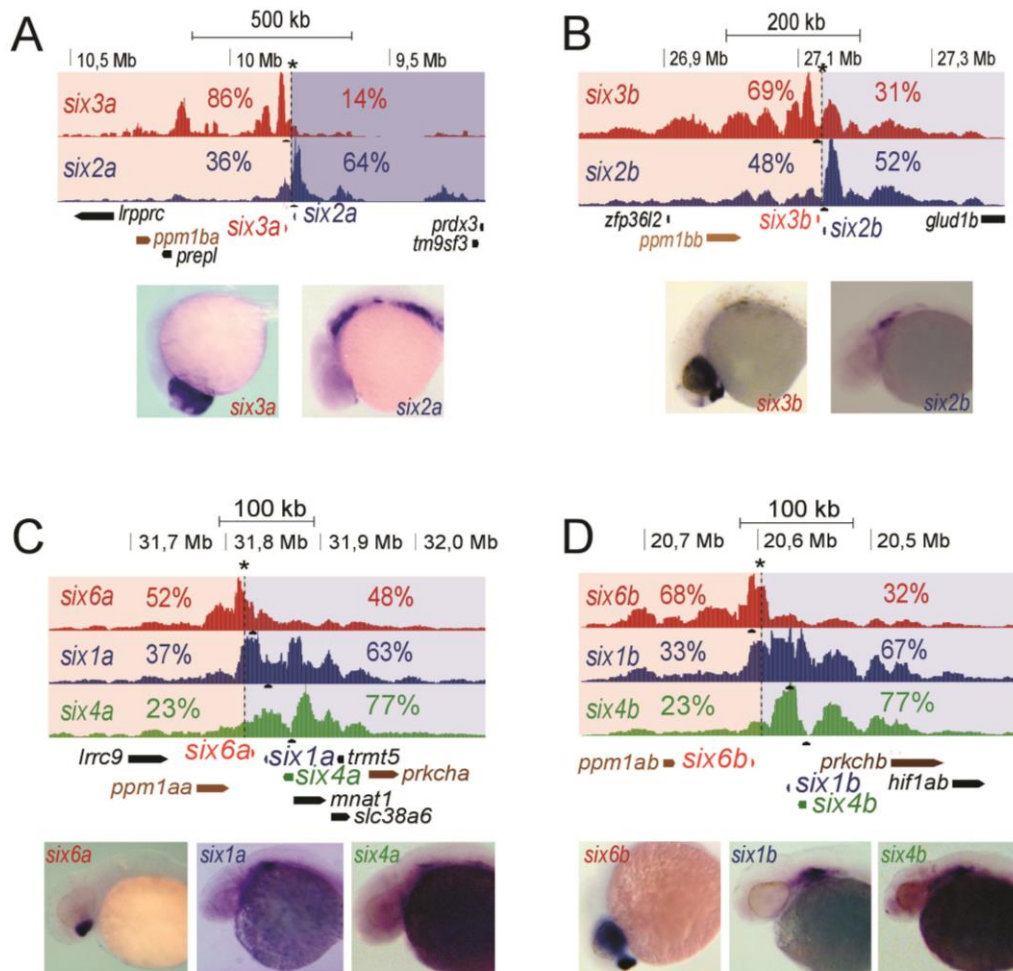


15 Fig RE_5

Fig. RE_5: Expresión diferencial de los genes *Six* de pez cebra presentes en un mismo complejo. Se muestra una representación esquemática de cada complejo *Six* analizado en pez cebra (la distancia entre los genes no es proporcional a la realidad). Debajo de la representación esquemática de cada complejo se muestra la expresión de cada gen en embriones a 48 hpf, excepto para *six6b* (asterisco negro) que no disponemos de In situ en ese estadio.

Usando embriones completos de pez cebra a 24hpf llevamos a cabo experimentos de 4C desde los promotores de cada uno de los genes *six* del resto de complejos (Fig. RE_6). La delimitación del entorno para el estudio se realiza de forma manual, mediante inspección visual de los datos y con el único criterio de incorporar la gran mayoría de los contactos significativos. Este criterio hace que la longitud de la región estudiada varíe de un complejo a otro (*six2a-six3a* → 1,3Mb, *six2b-six3b* → 1Mb, *six6a-six1a-six4a* → 460 kb, *six6b-six1b-six4b* → 400kb). Esta variación está íntimamente ligada con la longitud de los desiertos génicos que flanquean los genes del complejo. Así encontramos que el gen *ppm1ba* está a unas 460kb de *six3a*, mientras que su parálogo *ppm1aa* está a tan sólo 60kb de *six6a* (Fig. RE_6). Esta gran variación en la longitud de los entornos reguladores, fue la razón por la que la región a analizar se fijó específicamente para cada complejo en función de la extensión de la señal de 4Cs, mejor que considerar una longitud fija para todos. Para inferir un posible BTP (“borde topológico predicho”) en cada locus, realizamos el mismo análisis que para el caso del complejo *six2a-six3a*. Para el caso de los complejos de 3 genes calculamos el BTP comparando los genes 2 a 2. Tanto para el caso de los complejos *six6b-six1b-six4b* y *six6a-six1a-six4a*, observamos que el BTP es el mismo cuando usamos la pareja de genes *six6* y *six1* o la pareja *six6* y *six4* (Fig. RE_7, complejo 416b), tanto para el complejo a como para el b. Esto se deduce del hecho, de que el punto donde la diferencia en la distribución de la señal a ambos lados es máxima es el mismo (Fig. RE_7, asterisco y línea discontinua). Cuando comparamos la señal de las parejas *six4* y *six1*, en ambos complejos, observamos un comportamiento diferente. La señal acumulada de ambos genes (*six4a* y *six1a* en un caso, y *six4b* y *six1b* en el otro) discurre prácticamente solapada (Fig. RE_7, segundo panel), indicando una distribución de contactos muy similar entre ambos genes. Además cuando se busca el punto donde la diferencia en la distribución de la señal a ambos lados es máxima (gráfica de “diferencia de la señal acumulada” Fig. RE_7, tercer panel), existen dos máximos correspondientes a la región del promotor de cada gen (en vez de uno situado entre ambos genes) siendo además entre 3 y 4 veces de menor intensidad que cuando comparamos las parejas *six6* y *six1* o *six6* y *six4*. Esto sugiere que las parejas *six1-six4* no estarían separadas por ningún borde topológico, y que ambos genes pertenecerían al mismo TAD o dominio físico. Así pues en cada complejo de genes *Six* de pez cebra, tendríamos un único BTP (borde topológico predicho) que dividiría cada locus dos TADs. Un TAD albergaría en cada

caso al gen *six3* o *six6*, mientras que el TAD adyacente albergaría al gen o a los genes *six2* o *six1-six4*.



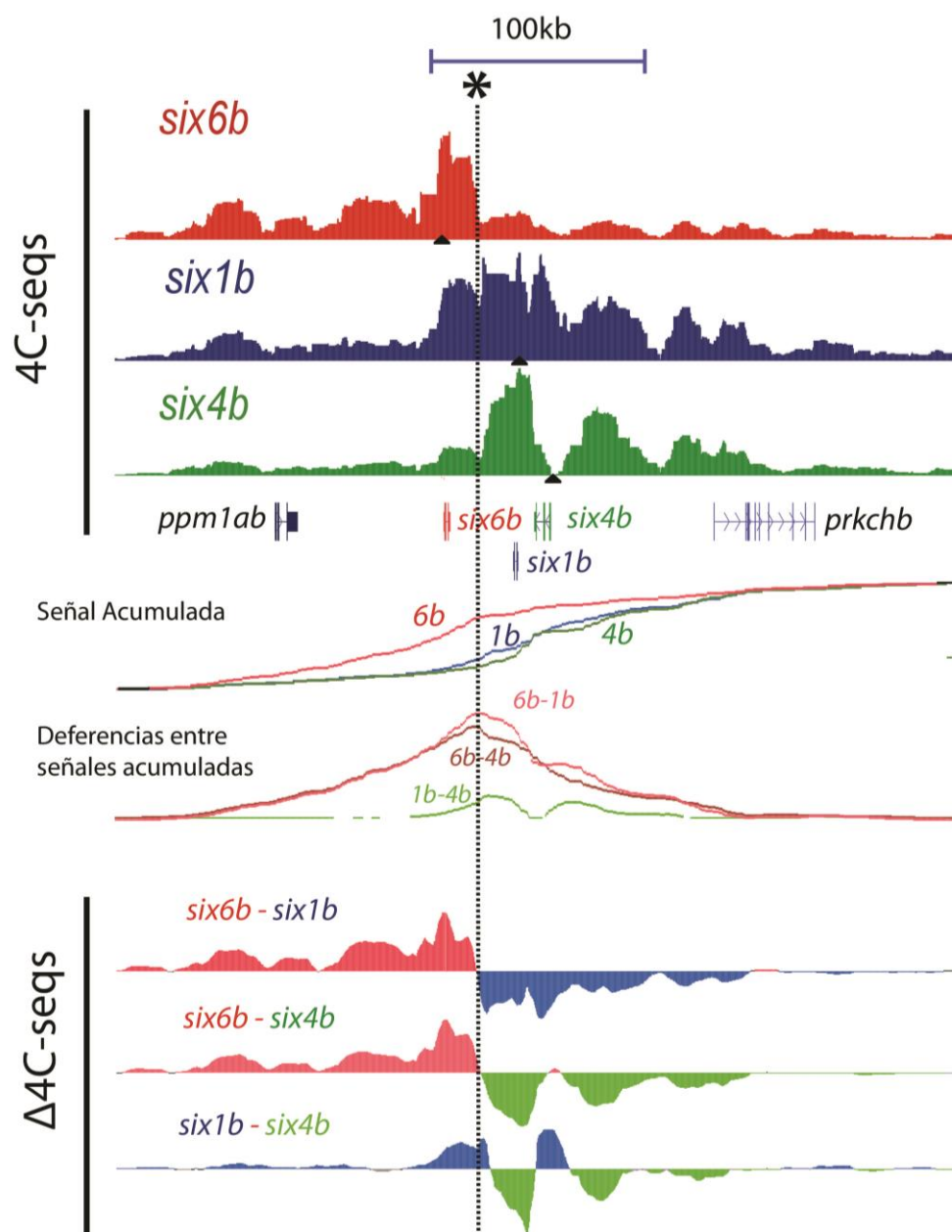
16 Fig RE_6

Fig. RE_6: Los cuatro complejos de genes Six de pez cebra se encuentran divididos en dos dominios físicos diferenciados. (A-D) Se muestran los perfiles de 4C-seq de los distintos genes *six3* realizados en embriones completos de pez cebra a 24hpf. Los puntos fijos están marcados con triángulos negros. Los bordes topológicos están marcados por un asterisco y una línea discontinua y los dos dominios físicos se encuentran sombreados en rojo y azul. Los porcentajes corresponden al porcentaje de señal de 4C a un lado y a otro del BTP. Además se muestra el patrón de expresión de cada gen a 24hpf. A) complejo *six2a-six3a*. B) complejo *six2b-six3b*. C) complejo *six6a-six1a-six4b*. D) complejo *six6b-six1b-six4b*.

Teniendo como referencia el BTP de cada complejo (Fig. RE_6, asteriscos) calculamos el porcentaje de señal de cada gen de un mismo complejo a cada lado del borde. En el caso del complejo *six3b-six2b* observamos que aguas arriba del BTP, *six3b* tiene el 69% del total de

la señal mientras que *six2b* sólo el 48%, siendo la señal aguas abajo del BTP de un 31% para *six3b* y de un 52% para *six2b*. Para los complejo *six6a-six1a-six4a* y *six6b-six1b-six4b* observamos también una distribución generalmente asimétrica de los contactos respecto del BTP. Así, en la región aguas arriba de los respectivos BTPs observamos una mayor frecuencia de contacto para los genes *six6a*(52%) y *six6b*(68%) que para los genes *six1a*(37%), *six1b*(33%), *six4a*(23%) y *six4b*(23%), ocurriendo lo opuesto aguas abajo de los BTPs: *six6a*(48%), *six6b*(32%), *six1a*(63%), *six1b*(67%), *six4a*(77%), *six4b*(77%). Por lo tanto observamos de nuevo como la mayor parte de los contactos de cada gen quedan restringidos al dominio físico al que el gen pertenece.

Hemos visto que a 24 hpf todos los complejos de genes *Six* de pez cebra, comparten el mismo principio organizativo tridimensional: la partición del complejo en dos dominios de asociación topológica (TADs) diferenciados. Al igual que en el caso del complejo *six3a-six2a* queremos saber si este principio organizativo es mantenido o cambia durante el proceso de desarrollo embrionario. Mediante el uso de embriones completos de pez cebra en distintos momentos del desarrollo, realizamos 4Cs desde los distintos promotores de los genes *Six*. Para los complejo *six3b-six2b* y *six6a-six1a-six4a* lo hicimos en los estadios de 80% de epibolia y 24 hpf. Para el complejo *six6b-six1b-six4b*, incluimos además el estadio de blástula (“Dome”) y 48hpf (Fig. RE_8). Podemos observar que no existen grandes cambios en el perfil de contacto de ninguno de los promotores, manteniéndose la distribución asimétrica respecto del BTP de complejo prácticamente inalterada. Además, los patrones de contactos de los genes situados en un mismo TAD poseen una distribución más similar que cuando comparamos el patrón de genes *Six* situados en TADs opuestos (Fig. RE_9, C segunda y tercera caja). Este fenómeno no puede explicarse como consecuencia de la distancia lineal entre promotores a un lado y a otro del borde topológico (Fig. RE_9, D). Si bien es cierto que la aproximación experimental con embriones completos, puede estar enmascarando la posible presencia de contactos dinámicos, los datos sugieren que la arquitectura global de cada complejo permanece constante independientemente del momento del desarrollo.



17 Fig RE_7

Fig. RE_7: Cálculo de Borde Topológico Predicho (BTP) en el complejo *six6b-six1b-six4b*. El panel superior muestra los 4C-seq realizados en embriones completos de pez cebra a 24hpf de los genes *six6b*, *six1b* y *six4b*. El segundo panel muestra la señal acumulada de 4C-seq para cada gen, mostrando el siguiente panel las diferencias entre dichas señales (6b-1b, 6b-4b y 1b-4b). El último panel muestra las diferencias entre las señales de 4C-seq de los distintos genes. El asterisco y línea discontinua marcan la localización del Borde Topológico Predicho.

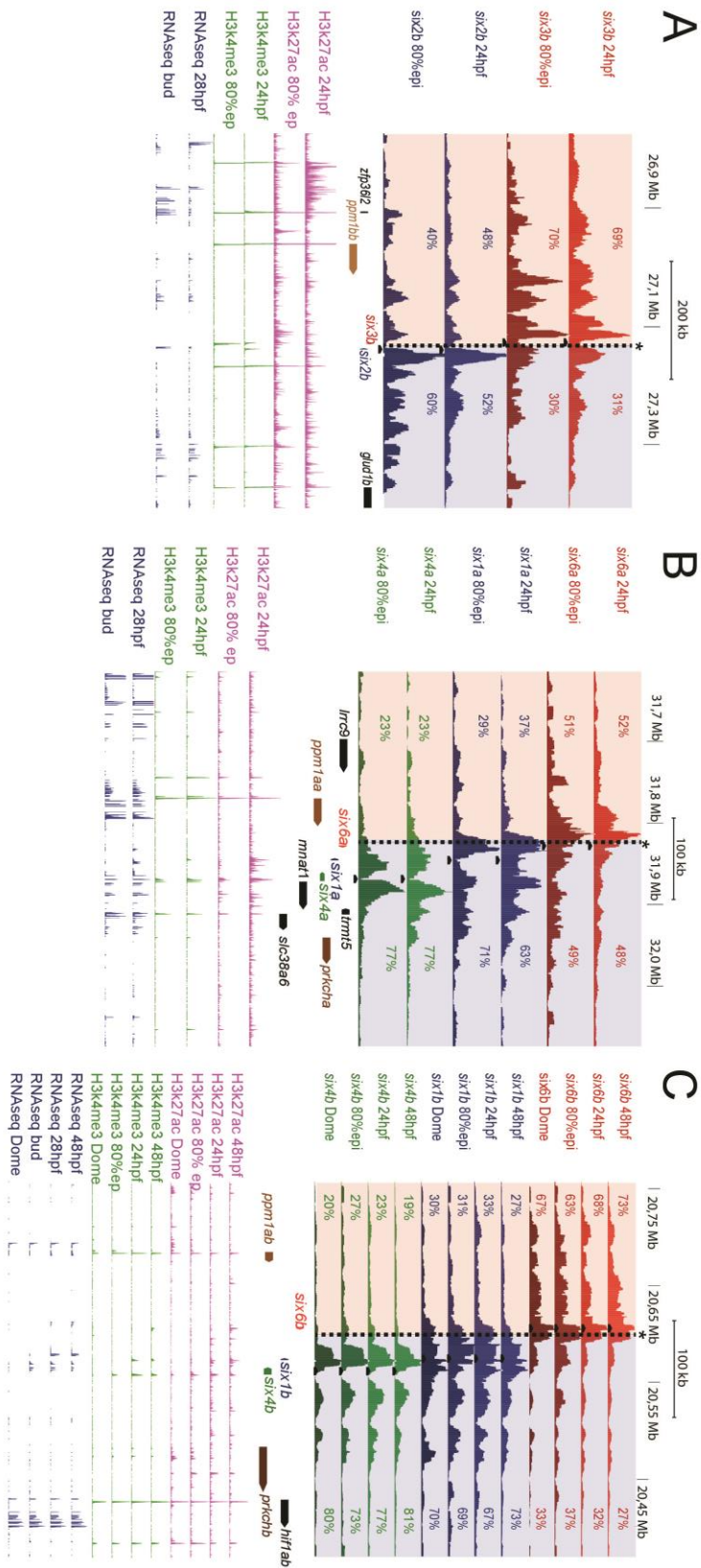


Fig. RE_8: Dinámica durante el desarrollo de los perfiles de contacto de 4C-seq, de las marcas de histonas H3K27ac y H3K4me3 y de ARN-seq en los complejos: six2b-six3b (A), six6a-six1a-six4a (B) y six6b-six1b-six4b (C). En todos los casos el BTP aparece marcado con un asterisco negro y una línea discontinua y los porcentajes corresponden al porcentaje de señal de 4C a un lado y a otro del BTP para cada promotor. Los 4C-seq se han realizado sobre embriones completos en los distintos estadios.

El análisis del patrón de contactos del resto de genes *six* de pez cebra, sugiere que el principio organizativo tridimensional encontrado en el locus *six2-six3a* se repite en el resto de complejos *Six* del pez cebra. Así cada complejo estaría dividido en dos TADs, que facilitaría la elaboración de los dos paisajes reguladores diferenciados asociados a cada complejo *Six* de pez cebra. Además como ocurriera en el complejo *six2a-six3a* este principio organizativo se mantendría constante durante el desarrollo embrionario e independiente del estado regulador de los genes *Six*.

4. ORGANIZACIÓN TRIDIMENSIONAL DE LOS COMPLEJOS *Six* EN RATÓN

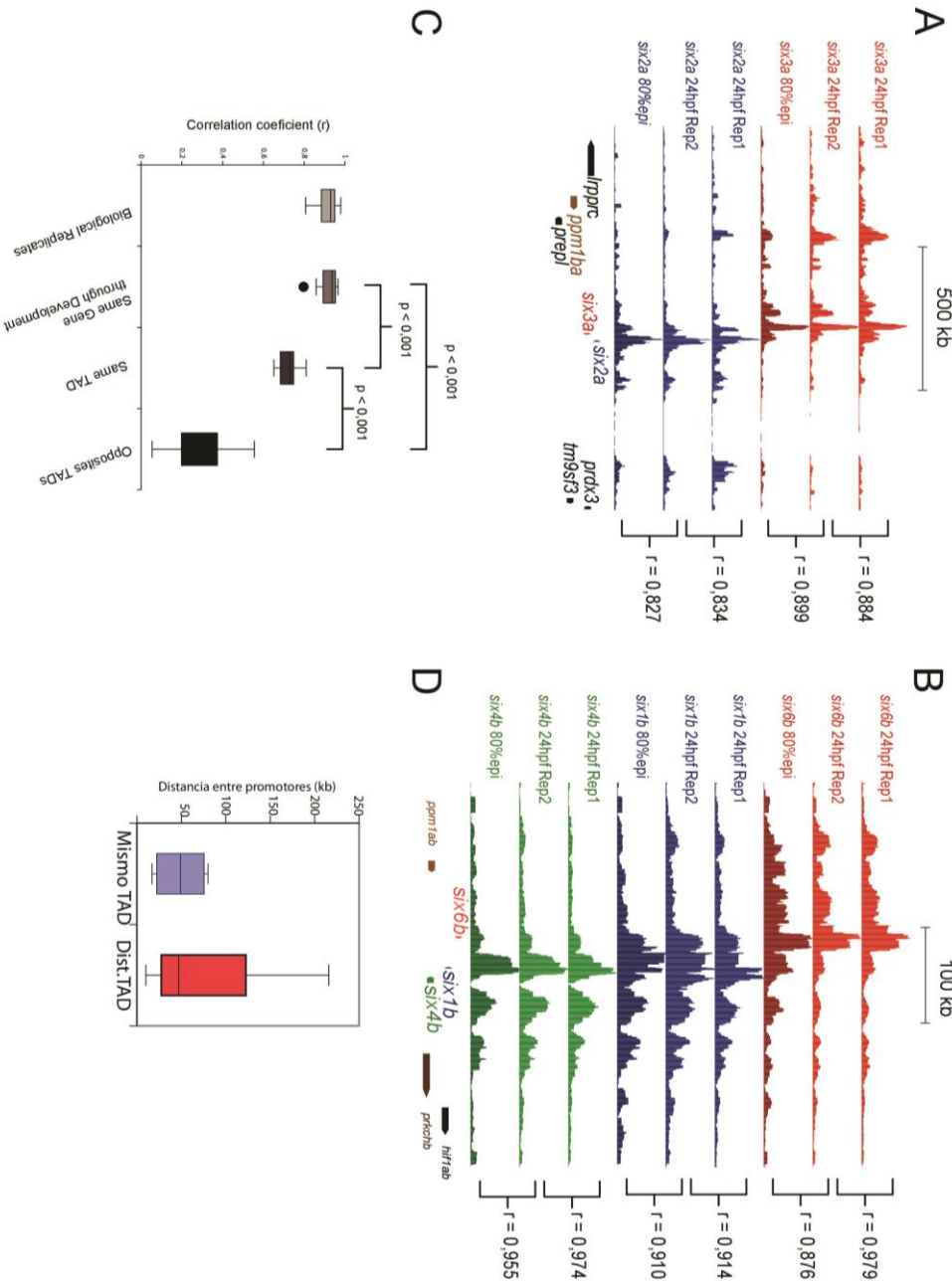
Hemos visto como todos los complejos de genes *Six* de pez cebra comparten una organización topológica similar que consiste en la división de cada complejo en dos TADs diferentes, que correlacionan con la generación de entornos reguladores separados que producen una activación transcripcional diferencial de los genes situados en uno u otro dominio físico. El siguiente paso fue comprobar si este escenario era específico de pez cebra o es compartido por otros vertebrados. Para tal fin, decidimos analizar el patrón de contactos de los genes *Six* agrupados en los dos complejos presentes en ratón: *Six3-Six2* y *Six6-Six1-Six4*. Usando embriones completos de ratón a 14.5 dpf realizamos 4C-seq desde los promotores de cada uno de los genes (Fig. RE_10, A y B). Al igual que para el caso del pez cebra, la delimitación de la extensión de la región a analizar se realizó de forma manual y específica para cada caso, teniendo como criterio que la región contuviera la gran mayoría de los contactos significativos de todos los genes integrados en el complejo. En esta ocasión la región a analizar para el complejo *Six3-Six2* abarca 1,4 Mb, longitud similar a la región analizada para el complejo *Six6-Six1-Six4* que es de 1,2 Mb. Ambas regiones analizadas abarcan los genes flanqueantes sinténicos: *Ppm1b* y *Ppm1a* aguas arriba de los complejos, y *Prkce* y *Prkch* aguas abajo.

Siguiendo la metodología utilizada en pez cebra, inferimos un único borde topológico (BTP) en cada complejo (Fig. RE_10, A y B). Usando este BTP como referencia determinamos, para cada complejo, el porcentaje total de contactos de cada promotor a ambos lados del borde. En el complejo *Six3-Six2* (Fig. RE_10, A), la mayoría de los contactos de *Six3* (67%)

se encontraban aguas arriba del BTP, mientras que en esta región sólo encontramos el 32% de los contactos de *Six2*. Aguas abajo del BTP la situación es la contraria, con un 33% de los contactos de *Six3* y un 68% de los contactos de *Six2*. En el complejo *Six6-Six1-Six4* tenemos una situación análoga (Fig. RE_10, B), encontrando aguas arriba del BTP un porcentaje de contactos mucho mayor para *Six6* (56%) que para *Six1* (19%) y para *Six4* (18%). La situación aguas abajo del BTP, en esta ocasión, sigue siendo la opuesta que aguas arriba, encontrando un porcentaje de contacto mucho mayor para *Six1* (81%) y *Six4* (82%) que para *Six6* (44%). Como ya ocurrió en pez cebra, estos datos sugieren una división de cada complejo en dos dominios físicos independientes, aislando los genes de cada complejo en dos entornos cromosómicos distintos. En concordancia con esta idea, vemos que los genes *Six* de un mismo complejo, situados en un mismo dominio físico, comparten muchos territorios de expresión, mientras que los genes de un mismo complejo pero localizados en distintos TADs poseen un patrón de expresión en gran medida diferente (Fig. RE_10, A y B paneles de la izquierda). Así por ejemplo tenemos que *Six1* y *Six4* están localizados en el mismo TAD y comparten expresión en los arcos branquiales, somitos o el mesénquima de las extremidades; sin embargo, *Six6* localizado en el otro TADs tiene unos territorios de expresión muy diferentes a la de los dos genes anteriores, como son el ojo o el cerebro anterior. En el caso del complejo *Six3-Six2*, donde cada gen pertenece a un dominio topológico distinto, los patrones de expresión son, a la vez, también muy diferentes. Así la expresión de *Six3* es similar a la de *Six6*, mientras que *Six2* es más parecido a la expresión de *Six1* y *Six4*, activándose en territorios como el mesénquima craneal, el epitelio nasal o los arcos branquiales.

Para corroborar la partición de los complejos de genes *Six* de ratón en dos dominios físicos, comparamos nuestros perfiles de contacto de 4Cs, con las matrices de contacto de HiC de células madre embrionarias de ratón disponibles en el buscador “WashU EpiGenome Browser” (<http://epigenomegateway.wustl.edu/>) (Fig. RE_10, A y B, paneles inferiores). A pesar de la diferencia en la aproximación experimental tanto técnica (4C frente HiC) como biológica (embriones completos a 14.5 dpf vs células madre embrionarias en cultivo), observamos que: 1) tanto el complejo *Six3-Six2* como el *Six6-Six1-Six4* se encuentran en la intersección de dos TADs, 2) La extensión de los perfiles de contactos obtenidos por 4C concuerdan de manera precisa con las extensión de los TADs, y 3) Los “bordes topológicos

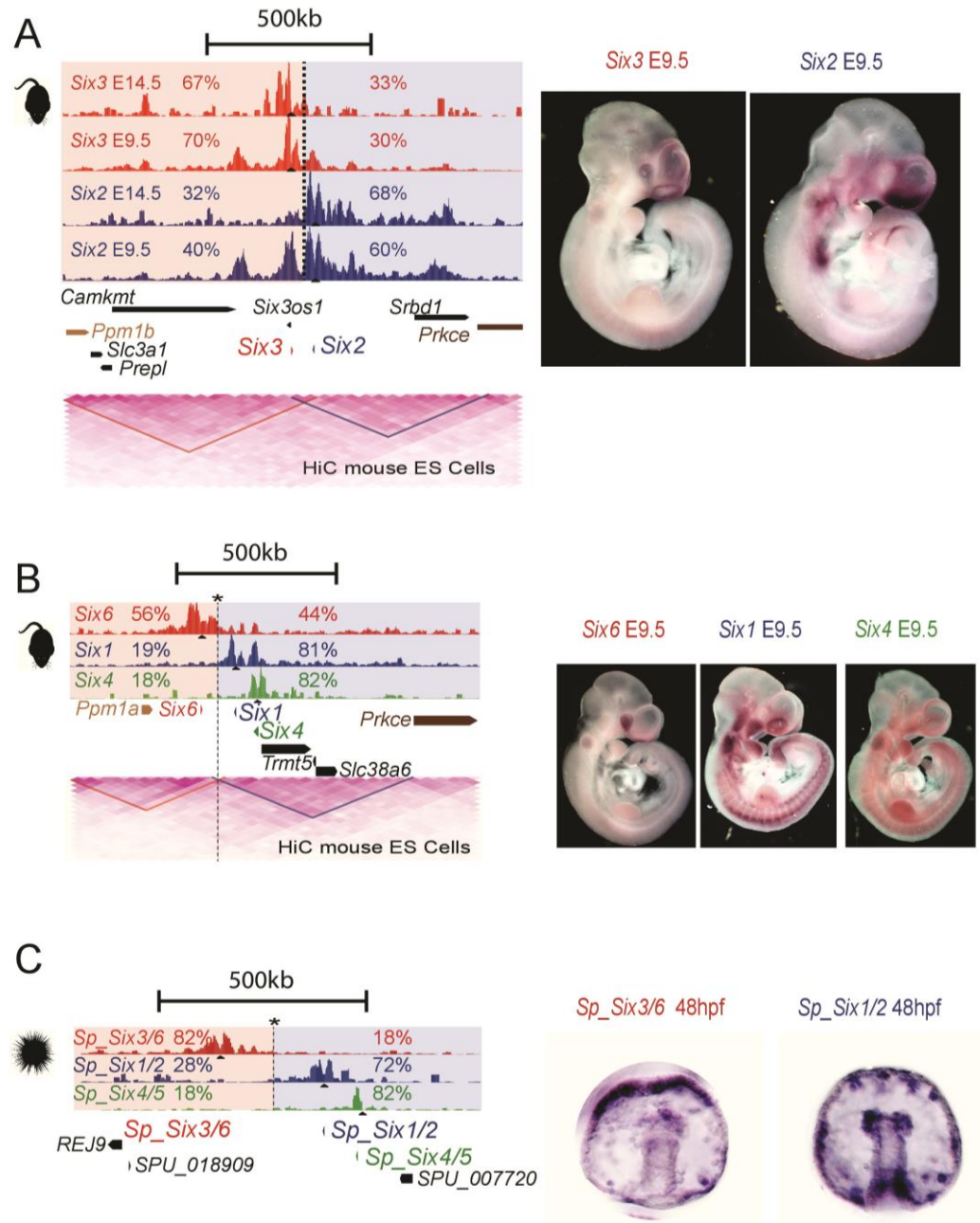
Fig. RE_9: Análisis de la similitud de los perfiles de 4C-seq entre réplicas biológicas, distintos momentos del desarrollo, y distintos dominios topológicos. A) Comparación del coeficiente de correlación de Pearson (p) entre dos réplicas biológicas (24hpf) y entre dos momentos del desarrollo (24hpf y 80%) de los perfiles de 4C-seq de los genes *six2a* y *six3a*. B) Lo mismo para los genes *six6b*, *six1* y *six4b*. C) Distribución y media de los coeficientes de correlación obtenidos al comparar todas las réplicas biológicas (incluyendo pez cebra, ratón y erizo de mar), todos los momentos de desarrollo, todos los genes de un mismo complejo presentes en un mismo dominio topológico, y todos los genes de un mismo complejo presente de dominios topológicos distintos. D) Distribución y media de las distancias entre los promotores de los genes *Six* presentes en un mismo complejo y localizados en el mismo dominio topológico y presentes en un mismo complejo pero localizados en distintos dominios topológicos (se han



19 Fig RE_9

predichos” por 4C coinciden con la posición del borde entre los dos TADs que conforman cada complejo. Estos dos conjuntos de datos independientes sugieren, con una alta confianza, que los complejos de genes *Six* en ratón siguen el mismo principio organizativo que los complejos de pez cebra.

Finalmente, al igual que en pez cebra, decidimos estudiar la dinámica de la organización tridimensional de los complejo *Six* en ratón en distintos momentos del desarrollo. Para ello utilizamos sólo el complejo *Six3-Six2*, y realizamos 4Cs desde ambos promotores en embriones completos a 9.5 dpf (Fig. RE_10, A). La predicción del BTP y el cálculo del porcentaje de contactos para cada gen a un lado y al otro de dicho borde, mostró una organización global muy similar en este estadio. Encontramos que en ambos estadios, aguas arriba del BTP tenemos alrededor de un 70% de los contactos de *Six3* mientras que para *Six2* tenemos alrededor del 40%. La distribución de los contactos aguas abajo del BTP es la opuesta, siendo de un 30% para *Six3* y un 60% para *Six2*. Como pasaba en pez cebra, la organización global del complejo *Six3-Six2* en dos dominios físicos diferenciados, parece ser independiente del estado de desarrollo, si bien, encontramos diferencias puntuales en los patrones de contacto de ambos genes.



20 Fig RE_20

Fig. RE_10: Los complejos Six de ratón y erizo de mar se encuentran divididos también en dos dominios topológicos. A) El panel superior muestra los perfiles de 4C-seq de los genes *Six3* y *Six2* obtenidos usando embriones completos a 9.5 y 14.5 dpf. El panel inferior muestra datos de HiC en células madre embrionarias de ratón, los dos TADs han sido delineados en rojo y azul. A la derecha se muestra el patrón de expresión de los genes *Six3* y *Six2* en embriones a 9.5 dpf. B) Lo mismo para los genes *Six6*, *Six1* y *Six4*. En esta ocasión los 4C-seq se han realizado sólo en a 14.5 dpf. C) A la izquierda, perfiles de 4C-seq de los genes *Six* de erizo de mar realizados sobre embriones completos a 48 hpf. A la derecha, patrón de expresión de los genes *Sp-Six3/6* y *Sp-Six1/2* en embriones a 48 hpf (el gen *Sp-Six4/5* no se expresa en este momento del desarrollo). En todos los casos junto a los perfiles de 4C-seq se muestran los BTPs calculados y el porcentaje de señal de 4C a un lado y al otro del mismo.

5. ORGANIZACIÓN TRIDIMENSIONAL DEL COMPLEJO *Six* EN ERIZO DE MAR

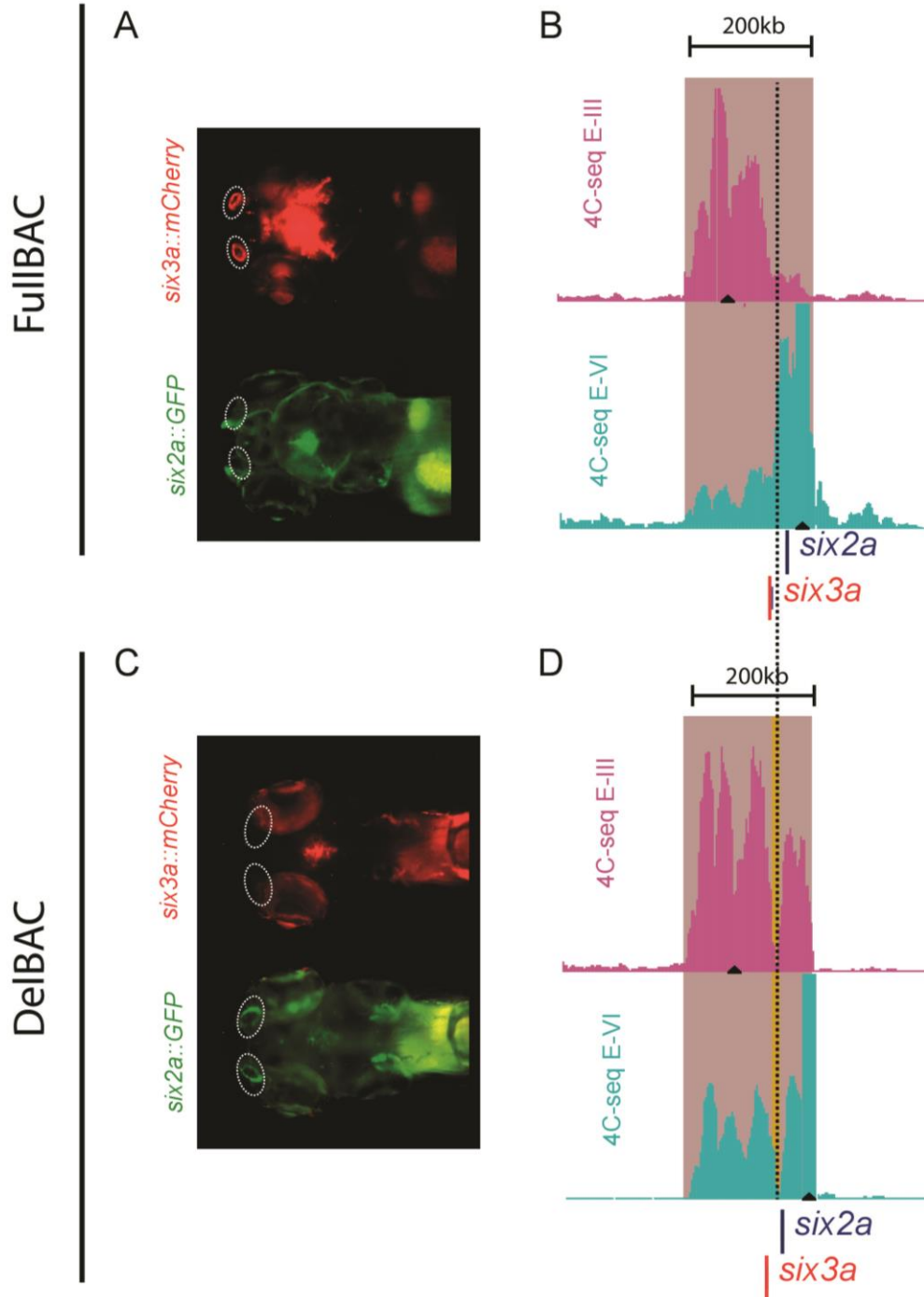
Hasta el momento hemos mostrado como los complejos de genes *Six* de pez cebra y ratón comparten un principio organizativo tridimensional común: la partición del complejo en dos dominios físicos aislados, que correlacionan con una expresión diferencial de los genes situados en cada dominio. Esto sugiere que este principio organizativo puede ser común a todos los vertebrados y que además precedería a la duplicación del complejo original de genes *Six*. Para comprobar esta hipótesis decidimos analizar la organización tridimensional del único complejo de genes *Six* presente en erizo de mar, *Strongylocentrotus purpuratus*, un deuterostomo invertebrado que se distanció de los vertebrados hace unos 540 millones años y fue el primer genoma deuterostomo no cordado que fue secuenciado. Posee un único complejo intacto de 3 genes *Six* organizados en tándem, homólogo a los de vertebrados, compuesto por los genes *Six3/6*, *Six1/2* y *Six4/5*. Durante la gastrulación, el gen *Six3/6* se expresa en territorios apicales neurogénicos, de forma muy distinta al gen *Six1/2* que se expresa en células mesénquimales y esqueletogénicas (Fig. RE_10 C, paneles de la izquierda), patrones equivalentes a los de vertebrados. El gen *Six4/5* no se expresa significativamente durante el proceso de gastrulación.

Usando embriones completos a 48 hpf (gastrulación tardía) realizamos experimentos de 4C desde los promotores de cada uno de los tres genes *Six* del complejo del erizo de mar. Para la localización de la posible región promotora de los genes, mapeamos las regiones codificantes de los genes que aparecen en la base de datos www.echinobase.org (antes www.SpBase.org). Para el gen *Six3/6* la secuencia SPU_018908.1, para el gen *Six1/2* la secuencia SPU_017379.1 y para el gen *Six4/5* la secuencia SPU_017380.1. Además nos ayudamos de los datos de ChIP-seq contra H3K4me3 generados en nuestro laboratorio, en embriones de *S. purpuratus* a 48 hpf. Para el análisis de los resultados seguimos los criterios usados con anterioridad en vertebrados (Fig. 10C). En esta ocasión analizamos una región de alrededor de 1Mb. Inferimos un único borde topológico (BTP) (Fig. RE_10, C, asterisco) situado entre los genes *Six3/6* y *Six1/2*. Aguas arriba del BTP, *Six3/6* presenta la mayor parte de sus contactos (82%) mientras que *Six1/2* y *Six4/5* sólo tienen una pequeña proporción de los mismos (28% y 18% respectivamente). Aguas abajo del BTP, tenemos la distribución

contraria con pocos contactos de *Six3/6* (18%) y la gran mayoría de la señal para *Six1/2* (72%) y *Six4/5* (82%). Estos datos sugerirían la posibilidad de que en erizo de mar el complejo de genes *Six* también estuviese distribuido en dos TADs adyacentes, contribuyendo al establecimiento de dos entornos reguladores diferenciados, uno para *Six3/6* y otro para *Six1/2* y *Six4/5*. Si bien en este caso hay que tener en cuenta que la distancia lineal entre el promotor de *Six3/6* y *Six1/2* es mucho mayor que en vertebrados, unas 240kb lo que supone 5 veces más que la media de la distancia de estos genes en los complejos de vertebrados (~50 kb). En consecuencia, la diferenciación entre la zona de contactos de *Six3/6* y *Six1/2* podría ser fruto exclusivamente de la distancia lineal entre los promotores y no de la existencia de dos dominios físicos diferenciados. No obstante, el decaimiento de la señal de 4C de *Six3/6* a la altura del BTP coincide justo con el comienzo de la señal de *Six1/2*, lo que sugiere una competición activa entre ambas señales. Además, la ausencia de señal de *Six1/2* justo aguas arriba del BTP no puede considerarse un simple efecto de la distancia lineal, ya que encontramos picos de señal unas 300kb aguas arriba del BTP cerca del gen *REJ9*.

6. CONTRIBUCIÓN FUNCIONAL DEL BORDE TOPOLÓGICO PRESENTE EN EL COMPLEJO *six3a-six2a* DE PEZ CEBRA.

Los resultados anteriores señalan a la conservación de un principio organizativo tridimensional que acompaña a la sintenia presente en los complejos *Six* de los deuterostomos. Este principio organizativo estaría basado en la división del complejo en dos dominios físicos o TADs adyacentes, separados por un borde topológico que ayudaría a limitar las interacciones físicas entre elementos reguladores de uno u otro TAD, favoreciendo la aparición de dos entornos reguladores diferenciados. La presión evolutiva habría favorecido el mantenimiento de este principio organizativo. Siguiendo esta hipótesis quisimos evaluar cuál era el papel funcional del borde topológico presente en estos complejos y cuál es su contribución en el control de la expresión de los genes *Six*. Para ello realizamos una aproximación experimental basada en el uso de cromosomas artificiales de bacterias (BACs), su modificación mediante recombinación homóloga en bacterias y su ensayo por transgénesis en pez cebra. La idea consistía en utilizar el BAC de pez cebra DKEY-74B2 que contiene la región que se extiende desde 120 kb aguas arriba de *six3a* hasta 47 kb aguas abajo de *six2a*. Esto significa que el BAC contiene ambas secuencias codificantes así como parte del paisaje regulador de ambos genes, y lo que es más importante, contiene el posible borde topológico presente entre ambos (Fig. RE_2_B). Mediante recombinación homóloga en bacterias reemplazamos las secuencias codificantes de los genes *six3a* y *six2a*, por los genes reporteros mCherry y GFP respectivamente. Además introducimos en él las secuencias de reconocimiento de la transposasa *tol2* necesarias para su transgénesis. Este BAC DK74B2-*six2a::GFP-six3a::mCherry-iTol2* (lo llamaremos FullBAC para simplificar) contiene la secuencia completa. Generamos también una versión modificada de este BAC, la cual consistió en la eliminación, mediante recombinación homóloga en bacterias, de la mayor parte de la región intergénica presente entre *six2a* y *six3a*. Concretamente eliminamos una región de 18kb que albergaría el hipotético borde topológico presente entre ambos genes (a esta nueva versión del BAC, $\Delta 18\text{kb-DK74B2-six2a::GFP-six3a::mCherry-iTol2}$, la llamaremos **$\Delta 18\text{BAC}$** , para simplificar). Estas 18kb eliminadas no presentan ninguna señal significativa en los datos de ChIP-seq de H3K27ac (elementos cis-reguladores activos) en distintos estadios de desarrollo de pez cebra generados en nuestro laboratorio, lo que descartaría la presencia de elementos reguladores en esta región.



21 Fig RE_11

Fig. RE_11: Papel funcional del borde topológico presente en el complejo *six3a-six2a*. Embriones transgénicos estables a 5 dpf portando las construcciones FullBAC (A) y $\Delta 18$ BAC (C), expresando GFP bajo el promotor de *six2a* y mCherry bajo el promotor de *six3a*. Nótese la transición en GFP-mCherry en la vesícula nasal entre las dos versiones del BAC (punteado blanco). B) Perfil de 4C-seq sobre embriones completos a 24hpf de pez cebra transgénicos transitorios inyectados con la versión FullBAC linealizada (Sfi-I), desde los elementos E-III y E-VI. D) Lo mismo que en (B) pero con embriones inyectados con la versión linealizada de $\Delta 18$ BAC. En (B-D) la región sombreada marca la extensión del BAC y los triángulos negros los puntos fijos. En (D) la línea amarilla marca la extensión de la región eliminada en la versión $\Delta 18$ BAC.

En resumen, teníamos dos versiones de un mismo BAC que contiene los genes *six2a* y *six3a* y parte de sus entornos reguladores, con la única diferencia de que una versión contendría el hipotético borde topológico existente entre ambos mientras que la otra no. Mediante el uso de la transposasa *tol2* procedimos a la transgénesis en pez cebra de ambas versiones de manera independiente. Tras la obtención de las líneas estables, analizamos el patrón de expresión de los reporteros incluidos en los BACs:

- **FullBAC:**

A 24hpf empezamos a observar la expresión de *six3a::mCherry* en el cerebro anterior y la retina ventral, mientras que no observamos una expresión clara *six2a::GFP* en ningún territorio (ANEXO 6 A, Founder 1). A 48hpf *six3a::mCherry* continuaba expresándose en el hipotálamo y en la retina ventral y comenzamos a ver una leve expresión de *six2a::GFP* en los arcos branquiales y levemente algunos puntos concretos del hipotálamo (ANEXO 6 B, Founder 1). A 5 dpf continuamos viendo una fuerte expresión de *six3a::mCherry* en hipotálamo y nervios craneales, una expresión más moderada y restringida en la retina ventral y aparece una nueva expresión en las vesículas olfatorias. Por su parte *six2a::GFP* presenta una leve expresión en distintas partes de la mandíbula, así como en la vesícula ótica y de una nueva en una región muy restringida del hipotálamo (Fig. RE_11, A, ANEXO 6 C Founder 1). Si bien más tarde obtuvimos otras dos líneas independientes con el mismo BAC, donde la expresión de *six2a::GFP* no producía ninguna señal significativa (ANEXO 6 A-C, Founders 2 y 3). La falta de mayor caracterización molecular de estas dos líneas, nos hace excluirlas de posteriores análisis.

- **Δ 18BAC:**

En esta versión del BAC no pudimos observar ninguna expresión ni para *six3a::mCherry* ni para *six2a::GFP* en las primeras 48hpf, pero tampoco podemos descartarla. A los 5dpf, en comparación con la expresión del FullBAC, *six3a::mCherry* muestra una reducción de señal en diversos territorios, así pierde prácticamente la expresión en la retina ventral, los nervios craneales y la vesícula olfatoria, quedando restringida a una región muy particular del hipotálamo (Fig. RE_11, C, ANEXO 6 C). Por otro lado, *six2a::GFP* tiene expresión en la vesícula ótica además de en la vesícula olfatoria, territorio que en la versión FullBAC pertenecía al dominio de expresión de *six3a::mCherry* (Fig. RE_11, A y C).

Estos datos de expresión muestran que la eliminación del hipotético borde topológico resultó en la pérdida de gran parte de los territorios de expresión de *six3a*, parte de los cuáles se incorporaron al dominio de expresión de *six2a*. Esto sugeriría que en ausencia del borde topológico, *six2a* es capaz de competir con *six3a* por elementos reguladores situados en el TAD opuesto. Para testar esta hipótesis, quisimos ver si existía algún cambio en los patrones de contactos de los elementos reguladores III y VI en ausencia del borde topológico. Anteriormente ya pudimos comprobar que el elemento III tenía actividad potenciadora que reproducía parcialmente el patrón de expresión de *six3a* y que el elemento VI hacía lo mismo con el patrón de *six2a* (Fig. RE_2_C). En consonancia con esto, cuando realizamos 4C-seq desde esos elementos, observamos que el elemento III contactaba fuertemente con *six3a* y su entorno regulador, mientras que los contactos más allá del borde topológico eran residuales (Fig. RE_2_B). Lo mismo ocurría con el elemento VI pero en esta caso con el promotor y el entorno regulador de *six2a*. En ausencia de una línea de pez cebra de $\Delta 18\text{BAC}$ estabilizada decidimos realizar los 4C-seq en líneas transitorias, F0. Para ello inyectamos las versiones FullBAC y $\Delta 18\text{BAC}$ linealizadas con la enzima Sfi-I (esta enzima corta en el “backbone” de los BACs, manteniendo intacta la organización de la región genómica). Pretendíamos que la linealización evitara contactos promiscuos como consecuencia de la naturaleza circular de los BACs. Así, fijamos los embriones inyectados con cada versión del BAC a 24hpf y realizamos el experimento de 4C-seq desde el elemento III y el VI. Estas PCRs no sólo mostrarán las interacciones de los elementos III y VI presentes en los BAC, sino también las interacciones endógenas. A pesar de esto, la señal que proviene de los BACs es cientos de veces más intensa que la endógena, como se desprende de los patrones de los 4C-seq donde la región abarcada por el BAC concentra la mayor parte de la señal total (Fig. RE_11, B y D). Esto significa que la contribución endógena de la señal es poco significativa. En el caso de la versión FullBAC, vemos como el patrón de contactos tanto del elemento III como del elemento VI respetan la frontera marcada por el borde topológico (Fig. RE_11, B), tal y como ocurría con los patrones de contacto de los elementos endógenos. Sin embargo, en el caso de la versión $\Delta 18\text{BAC}$, tanto los contactos del elemento III como del elemento VI son capaces de contactar más allá del borde topológico con una frecuencia similar a la que contactan con su entorno natural (Fig. RE_11, C). Si bien es necesario la obtención de repeticiones biológicas del experimento estos datos sugieren que la ausencia del hipotético borde topológico, disipa en cierta medida la constricción física que mantiene aislados los elementos reguladores de uno y otro TAD, permitiendo el establecimiento de contactos entre elementos de ambas regiones.



22 Fig RE_12

Fig. RE_12: La presencia de un nuevo entorno genómico no genera nuevos contactos significativos de forma automática. A) Perfil de 4C-seq en embriones completos transgénicos FullBAC a 24 hpf realizados sobre la GFP y la mCherry. A la izquierda se muestran los contactos con la región endógena del BAC, a la derecha se muestran los contactos con la región circundante donde el BAC ha sido insertado. Las barras negras simbolizan secuencias codificantes. **B)** Los mismo que en A) pero usando otra línea transgénica independiente de la misma versión FullBAC.

Sin duda el análisis mediante 4C en líneas estables de ambas versiones del BAC ayudaría a describir con mayor rigor las consecuencias estructurales de la eliminación del borde topológico. Recientemente hemos conseguido estabilizar varias líneas de la versión “FullBAC”(que contiene el borde topológico). Sobre 2 de estas líneas hemos realizado 4C-seq desde el gen GFP (punto fijo correspondiente a *six2a*) y desde el gen mCherry (punto fijo correspondiente a *six3a*). A falta de analizar los datos de forma cuantitativa, los perfiles de contactos de estos nuevos 4C se asemejan cualitativamente a los obtenidos desde los promotores endógenos de los genes *six2a* y *six3a*. Por otro lado, no sólo somos capaces de capturar contactos con los elementos internos del BAC sino que también capturamos contactos con elementos exógenos correspondientes a los sitios de inserción de los BACs

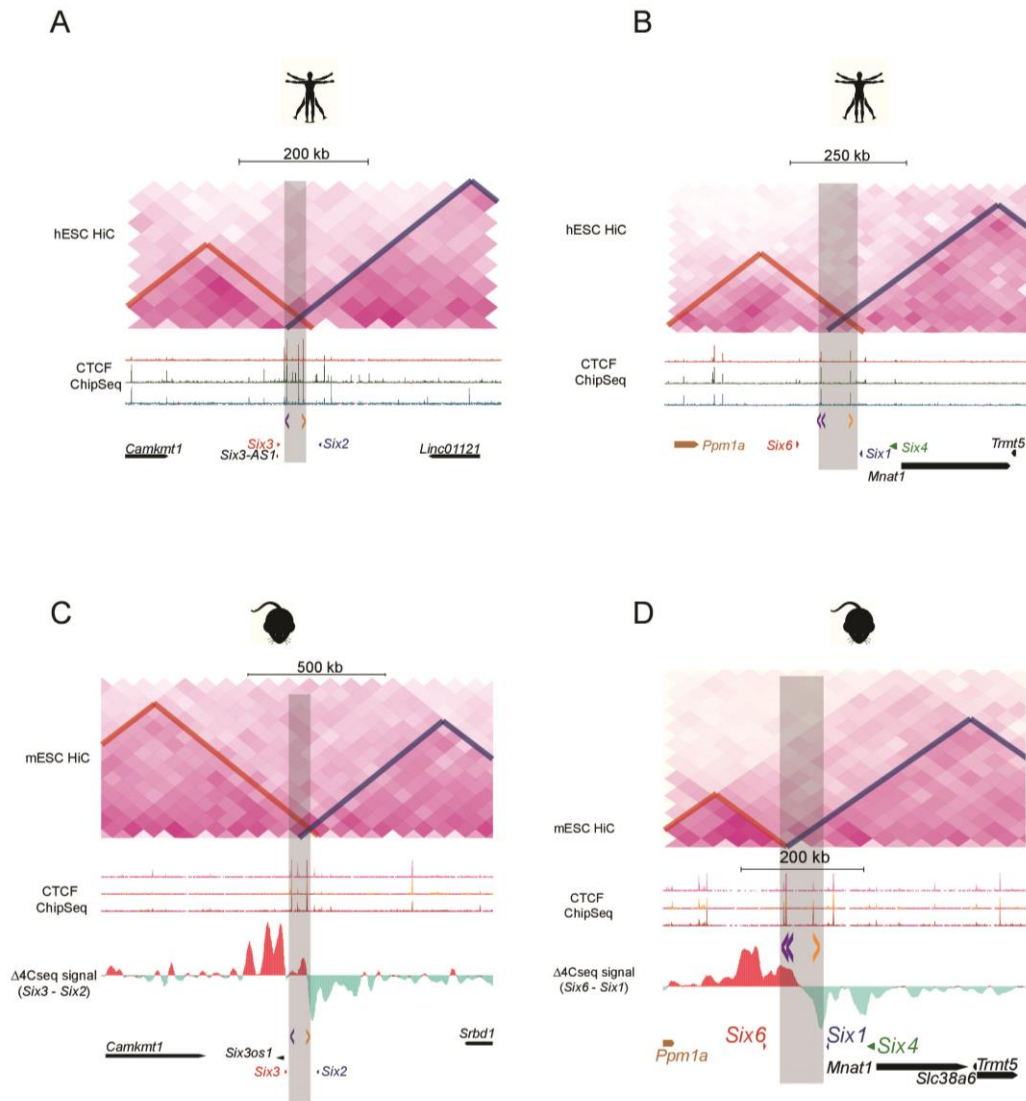
(Fig. RE_12). Destacar que pese a producirse, los contactos exógenos que encontramos, son de una intensidad significativamente menor que los del interior del BACs. Esto sugiere que pese a la adquisición por parte del complejo *six3a-six2a* de un nuevo entorno genómico, éste entorno y los elementos funcionales que pudiera contener no se han incorporado de manera automática a los paisajes reguladores de ambos genes.

7. BÚSQUEDA DE LOS DETERMINANTES DE LA ORGANIZACIÓN TRIDIMENSIONAL DE LOS COMPLEJOS *Six*.

Hasta aquí hemos mostrado como el análisis de la organización tridimensional de los complejos *Six* sugiere la posibilidad de que exista un principio organizativo común evolutivamente conservado en dichos complejos en deuterostomos, cuya principal característica sería la presencia de un borde topológico entre los genes *Six3/6* y los genes *Six1/2-4/5*, necesario para la prevención de la competencia de los promotores de dichos genes por elementos reguladores situados a lado opuesto del borde. Así que ahora quisimos explorar cuáles podrían ser los posibles determinantes moleculares responsables de esta partición topológica de los complejos *Six*.

Buscamos marcas asociadas a los bordes de los TADs, que estuvieran conservadas en todos los bordes topológicos de los complejos *Six* de los deuterostomos analizados. El estudio de la organización de los complejo *Six* de pez cebra durante el desarrollo embrionario, reveló que dicha organización aparecía independiente del estado regulador de los genes del complejo, este indicio nos hizo centrar la búsqueda de los determinantes moleculares de la arquitectura de los complejos *Six* en proteínas arquitectónicas de expresión ubicua más que en factores de transcripción de expresión tejido específica. La proteína de unión a ADN, factor de unión CCCTC (CTCF), se ha propuesto como instrumental para la configuración tridimensional de la cromatina (Rao y cols 2014). Además CTCF está enriquecido en los bordes de los TADs (Hou y cols 2012) y facilita bucles de ADN entre sitios de unión de CTCF, en los cuales estas proteínas quedan orientadas de forma convergente (cabeza-a-cabeza). Por estas razones decidimos explorar la posibilidad de que CTCF tuviese un papel relevante en la formación del borde topológico presente en los complejos *Six* de los deuterostomos.

Para empezar inspeccionamos los complejos *Six* en humanos, donde también tenemos dos complejos *Six*, el compuesto por los genes *SIX3* y *SIX2*, y el compuesto por los genes *SIX6*, *SIX1* y *SIX4*, organizados de la misma manera que los complejo *Six* de ratón. Los datos obtenidos mediante ChIP-seq de CTCF en distintos tipos celulares revelaron un enriquecimiento de la señal en la intersección de los dos TADs (Fig. RE_13, A y B) que albergan a cada complejo *Six*. En esta intersección existen dos picos de señal de unión de CTCF fuertes presentes en todos los tipos celulares (Fig. RE_13 A y B, flechas moradas y amarillas) que, a su vez, están envueltos en contactos de larga distancia hacia direcciones opuestas, como revelan datos públicos de ChIA-PET para CTCF. Estos dos sitios de unión para CTCF se encuentran orientados de manera divergente el uno respecto del otro (cola a cola), mientras que cada uno se encuentra orientado de forma convergente (cabeza a cabeza) con el sitio con el que establecen el contacto de larga distancia obtenido mediante ChIA-PET. Estos dos sitios de unión a CTCF orientados de manera divergente que marcan el borde entre los dos TADs que albergan cada complejo *Six* de humano, se encuentran conservados en secuencia en mamíferos. Por esta razón, buscamos si dichas secuencias en ratón se encuentran también ocupadas por CTCF orientados de manera divergente. Inspeccionamos los datos de ChIP-seq de CTCF en distintos tipos celulares de ratón. Observamos que tanto en el complejo *Six3-Six2* como en el complejo *Six6-Six1-Six4* de ratón, dichas secuencias conservadas también se encuentran ocupadas en todos los tipos celulares por las proteínas CTCF orientadas de manera divergente. En ratón, la posición marcada por sitios de unión a CTCF orientados de manera divergente marca también la intersección entre los dos TADs (Fig. RE_13, C y D) que componen cada complejo *Six*, que coincide a su vez con el borde topológico predicho (BTP) a partir de los datos de 4Cs obtenidos por nosotros.

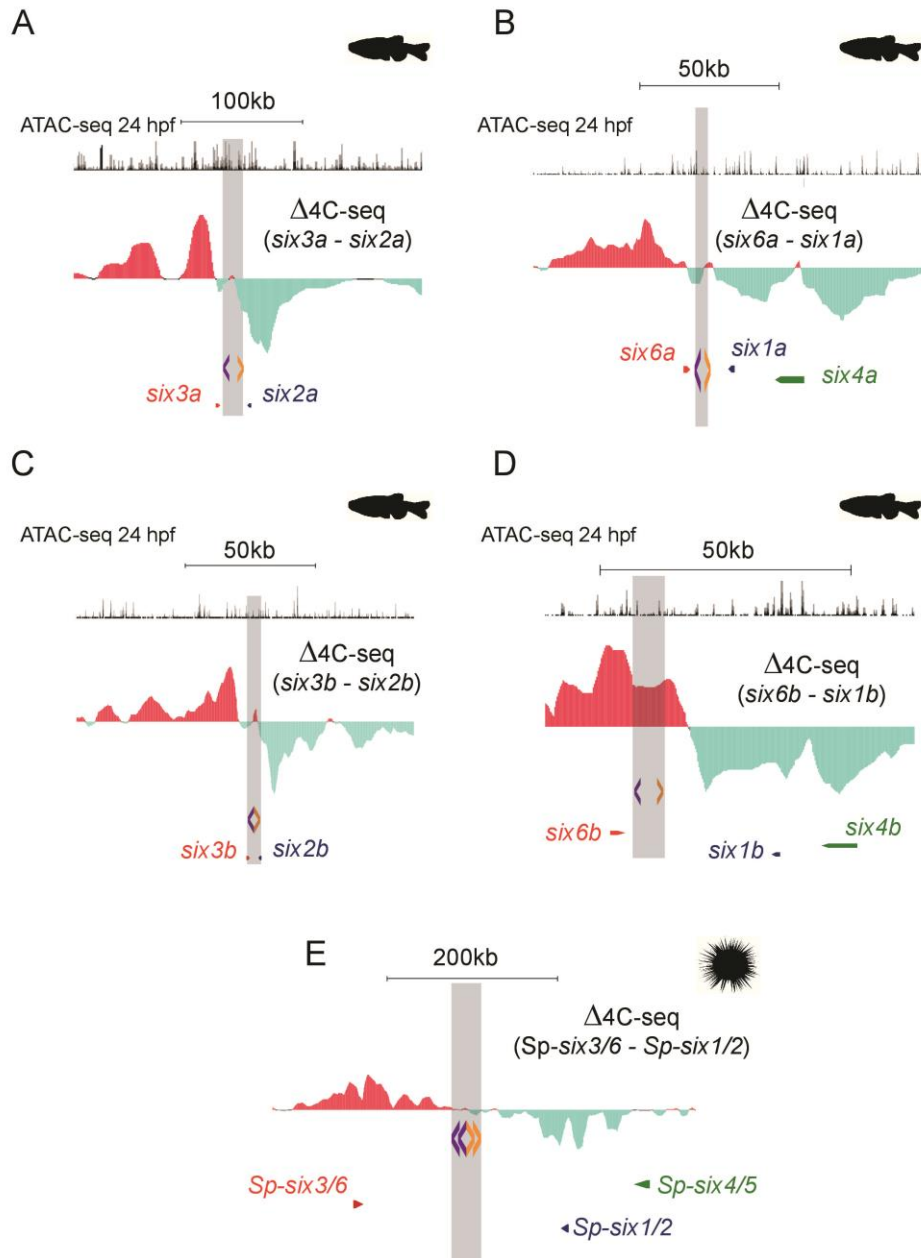


23 Fig RE_13

Fig. RE_13: La presencia de sitios de unión de CTCF orientados de manera divergente caracteriza el borde topológico presente en los complejos Six de humano y ratón. A y B) Muestran de arriba abajo: datos de HiC en células madre embrionaria humanas con los dos TADs marcados en rojo y azul, la distribución genómica de sitios de unión a CTCF en 3 tipos celulares distintos, y la orientación de los sitios de unión a CTCF constitutivos presentes en el borde topológico marcada por una cabeza de flecha morada (significa que el sitio en la cadena negativa, orientado con la cabeza hacia la izq.) y una cabeza de flecha amarilla (significa que el sitio de unión está en la cadena positiva, orientado hacia la derecha) tanto para el complejo *SIX2-SIX3* (A) como para el complejo *SIX4-SIX1-SIX6* (B). C y D) Los mismo que (A) y (B) pero para los complejo *Six2-Six3* (C) y *Six4-Six1-Six6* (D) de ratón, en ambos se incluye un panel representando la diferencia de la señal de 4C entre los genes *Six3/6* y los genes *Six2/1*. En todos los paneles la región sombreada en gris representa a aquella región marcada por los sitios de unión a CTCF divergentes presentes en el borde topológico.

El siguiente paso consistía en determinar si esta configuración de sitios de unión divergentes de CTCF también estaba presente en los bordes topológicos predichos (BTPs) para los cuatro

complejos *Six* de pez cebra. No disponíamos de datos de ChIP-seq de CTCF en pez cebra, por lo que utilizamos datos de ATAC-seq generados en nuestro laboratorio usando embriones completos de pez cebra a 24 hpf (Gehrke y cols 2015). La técnica ATAC-seq permite la identificación de regiones de cromatina abierta con una muy alta resolución, permitiendo determinar sitios de unión de factores de unión al ADN. En la región de los BTPs de cada complejo *Six* de pez cebra, determinada por la diferencia entre la señal de los 4Cs de los genes *Six* situados a un lado y a otro de dicho borde (Fig. RE_14, A-D), buscamos picos de ATAC-seq que contuvieran sitios de unión a CTCF. Para esto usamos la herramienta de búsqueda de sitios de unión de ADN de la base de datos JASPAR, y seleccionando el sitio de unión de CTCF de vertebrados de la misma. Para determinar un umbral de valor o puntuación que nos permitiera considerar un resultado como positivo, analizamos sólo las secuencias de humano y ratón que sabemos que se unen a CTCF presentes en los bordes topológicos de los complejos *Six*. La mayor puntuación obtenida es de 23 para la región de humano más cercana al gen *SIX1* (hCTCFsIX1) y la menor fue de 12 para la región de ratón mCTCFsIX6(2), (tabla RE_1). Por tanto en el caso del pez cebra, decidimos considerar positivos aquellos picos de ATAC-seq, donde la herramienta de búsqueda de matrices de peso por posición de la base de datos JASPAR predecía la presencia de una matriz de peso para CTCF con un valor igual o superior a 12. Con estas condiciones, encontramos dos sitios de unión a CTCF en la región correspondiente al borde topológico predicho de cada complejo *Six* de pez cebra (Fig. RE_14, A-D). Este par de sitios de unión a CTCF que marcan cada borde topológico de los complejo de pez cebra se encuentran orientados de forma divergente (cola a cola) uno respecto del otro, como pasa con los sitios de ratón y humano. El detalle de la localización de la matriz CTCF dentro de los picos de ATAC-seq se puede ver en la Fig. RE_15. Tan sólo el sitio de unión a CTCF más cercano a *six1a*, zCTCFsIX1a, no cumple el criterio del “score” de JASPAR que en este caso es de 11 (tabla RE_1). Estos datos sugieren que esta configuración de sitios de unión de CTCF orientados de manera divergente, es una marca común en la intersección de los TADs que albergan los complejos *Six* de vertebrados. Quisimos ver ahora, si esta marca estaba también presente más allá de vertebrados. Para esto tratamos de escanear sitios de unión a CTCF en la región del borde topológico predicho del complejo *Six* de erizo de mar. En esta ocasión no contamos con datos de ChIP-seq o ATAC-seq que nos permitieron delimitar experimentalmente las regiones a analizar. Así pues seleccionamos una de región que iba desde 10kb aguas arriba del BTP hasta 10kb aguas abajo del mismo.



24 Fig RE_14

Fig. RE_14: La presencia de sitios de unión de CTCF orientados de manera divergente caracteriza el BORDE TOPOLÓGICO PREDICHO presente en los complejos Six de pez cebra y erizo de mar. **A-D)** Muestran de arriba abajo: datos de ATAC-seq obtenidos con embriones completos pez cebra a 24hpf, diferencia de la señal de 4C-seq entre los genes *six3/6* y *six1/2* en embriones a 24hpf y por último la orientación de los sitios de unión a CTCF localizados en regiones positivas para ATAC-seq mediante JASPAR, flecha morada (cadena negativa) y flecha amarilla (cadena positiva), para los cuatro complejos *Six* de pez cebra: **A)** complejo *six2a-six3a*. **B)** complejo *six2b-six3b*. **C)** complejo *six6a-six1a-six4b*. **D)** complejo *six6b-six1b-six4b*. **E)** Muestra la diferencia entre la señal de 4C-seq de los genes *Sp-Six3/6* y *Sp-Six1/2* de erizo de mar, debajo se muestra la orientación de los sitios de unión a CTCF localizados mediante JASPAR en la región de alrededor del BTP.

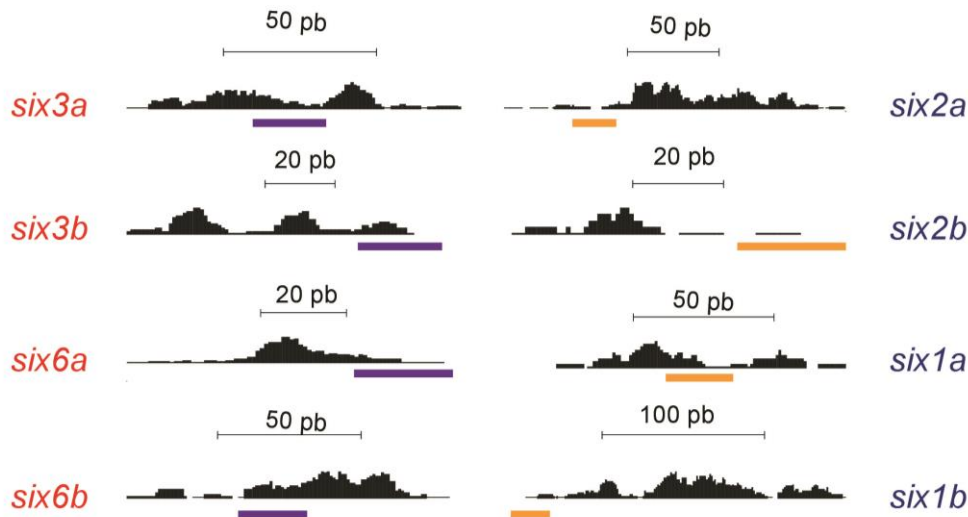
Resultados

Escaneamos esta región con JASPAR, usando la matriz de CTCF de vertebrados y encontramos cuatro resultados para CTCF (Fig. RE_14, D, tabla RE_1), dos aguas del BTP orientados con la cabeza aguas arriba, y dos resultados aguas abajo del BTP orientados con la cabeza aguas abajo. Si bien 3 de estos resultados (SpCTCFsix3/6(1), SpCTCFsix1/2(1) y SpCTCFsix1/2(2)) no alcanzaban la puntuación umbral de 12 marcada para pez cebra, es razonable pensar que debido a la distancia evolutiva de vertebrados con erizo de mar es probable que la matriz de unión para CTCF presente también más diferencias, a pesar de lo cual JASPAR es capaz aún de encontrar resultados con “scores” significativos (“8” para SpCTCFsix3/6(1), “11” para SpCTCFsix1/2(1) y “10” para SpCTCFsix1/2(2))(tabla RE_1). Siendo los resultados en este caso, de un menor soporte experimental que en vertebrados, irían en línea con la posibilidad de que los mecanismos moleculares que se encuentran tras la organización tridimensional común de los complejo *Six* en ratón, pez cebra y erizo de mar; sean también compartidos y tengan como agente principal la disposición divergente de sitios de unión a CTCF.

Nombre del element CTCF	Especie	Gen circundante	Coordenadas de la matriz CTCF			MATRIZ	Valor
			Cromosoma	Inicio	Fin		
hCTCFsix6(1)	Human	Six6	chr14	61026278	61026302	TGGACAGCAGGGGGCTCTC	16
hCTCFsix6(2)	Human	Six6	chr14	61029012	61029036	CCGCCACTTGGAGGCAGTG	13
hCTCFsix1	Human	Six1	chr14	61092245	61092270	TGGCCAGCAGGTGGCACTC	23
hCTCFsix2	Human	Six2	chr2	45182932	45182961	TCGCCGCGAGGTGGCAGCA	16
hCTCFsix3	Human	Six3	chr2	45208915	45208940	TCTCCAGCAGGTGGCGCCA	21
mCTCFsix6(1)	Mouse	Six6	chr12	74074290	74074314	ATGCCACCAGGGGGCTCTC	17
mCTCFsix6(2)	Mouse	Six6	chr12	74077887	74077911	TGTACGCCAGGTGGTGCTG	12
mCTCFsix1	Mouse	Six1	chr12	74122104	74122128	TATCCAGCAGGGGGCACTC	20
mCTCFsix2	Mouse	Six2	chr17	86033160	86033184	CGGCCGCGAGGTGGCAGCA	16
mCTCFsix3	Mouse	Six3	chr17	86064479	86064503	TCTCCAGCAGGTGGAGCCA	18
zCTCFsix6a	Zebrafish	six6a	chr13	31832136	31832159	TGACCACCAGAGGTTCGCAA	19
zCTCFsix1a	Zebrafish	six1a	chr13	31835564	31835588	TTGCCACCAGAGGGCAGAA	11
zCTCFsix6b	Zebrafish	six6b	chr20	20600656	20600680	CTAACTGAAGAGGGCGCTC	12
zCTCFsix1b	Zebrafish	six1b	chr20	20596398	20596422	TGTCCACTAGAGGGAACCA	19
zCTCFsix2a	Zebrafish	six2a	chr13	9819167	9819191	TTCCCAACAAGATGGCGTAA	13
zCTCFsix3a	Zebrafish	six3a	chr13	9804005	9804029	CGTTCAGGAGAGGGCGCCA	14
zCTCFsix2b	Zebrafish	six2b	chr12	27133491	27133515	TCACCACATGATGGCGACA	12
zCTCFsix3b	Zebrafish	six3b	chr12	27134370	27134394	TGGCCAGCAGAGGGTGCTT	18
SpCTCFsix3/6 (1)	Sea Urchin	Sp_six3/6	Scaffold143	896688	896712	TGACCAACAGACAGCGGCC	8
SpCTCFsix3/6 (2)	Sea Urchin	Sp_six3/6	Scaffold143	892778	892812	CGGCCAGGAGATGGAGAAG	13
SpCTCFsix1/2 (1)	Sea Urchin	Sp_six1/2	Scaffold143	888274	888298	CGGCCAGGAGATGAAGCAG	11
SpCTCFsix1/2(2)	Sea Urchin	Sp_six1/2	Scaffold143	883271	883297	ATGACGACAGAGGGCGGCA	10

Tabla RE_1

Tabla RE_1: Matrices CTCF detectadas mediante JASPAR en la señales de ATAC-seq presentes en los bordes topológicos de los complejos Six de humano, ratón, pez cebra y erizo.

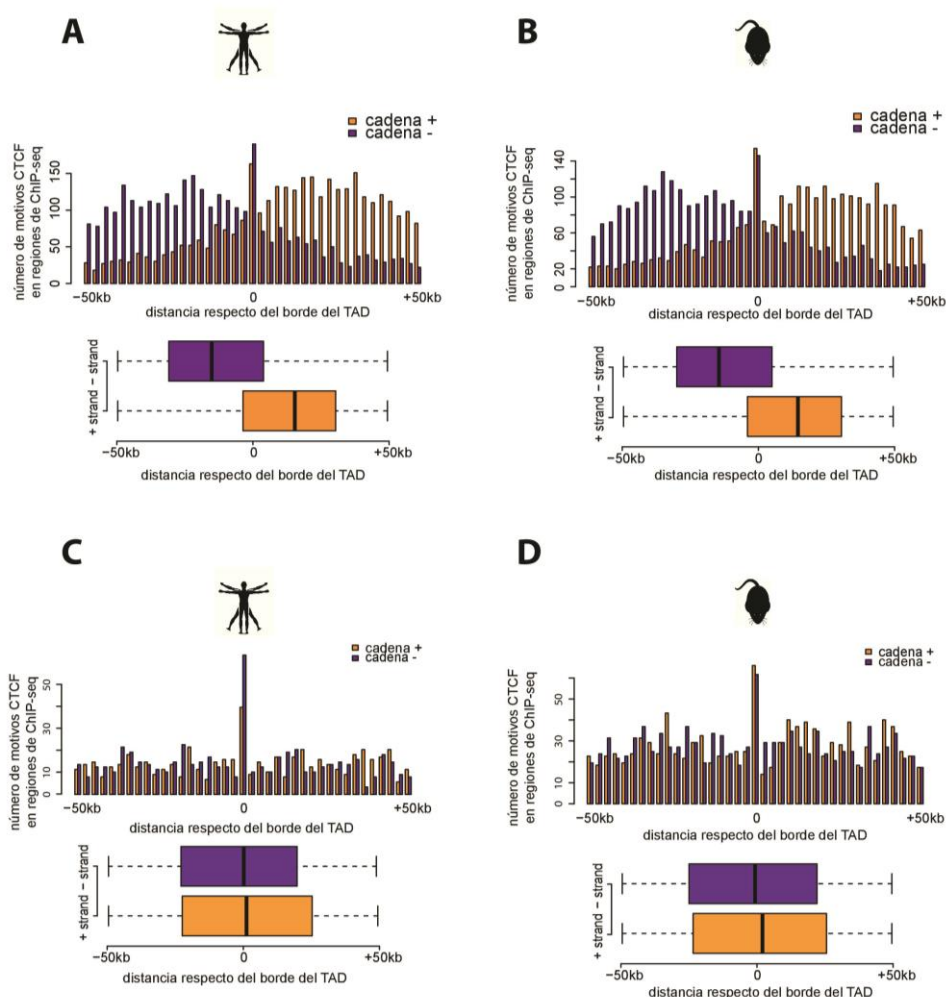


25 Fig RE 15

Fig. RE_15: Detalle de los sitios de unión a CTCF localizados en las regiones positivas de ATAC-seq en los bordes topológicos de los complejos *Six* de pez cebra. Se muestran los 8 picos de ATAC-seq (en negro) bajo los cuales se localiza la posición exacta de la matriz de unión a CTCF (barras moradas o amarillas) revelada por JASPAR. A izquierda y derecha aparecen los genes más cercanos.

Los datos mostrados sugieren que, en parte, el borde topológico presente entre los TADs que albergan los complejo *Six* de ratón, pez cebra y erizo de mar están determinados por la presencia de sitios de unión a CTCF orientados de manera divergente. Así que quisimos comprobar hasta qué punto esta marca molecular es una característica general de los bordes de los TADs. Para este fin usamos los datos disponibles de HiC en células madre embrionarias de ratón y humano. Analizamos una región que iba desde 50kb aguas arriba del borde de un TAD hasta 50kb aguas debajo de dicha posición para todo los bordes de los TADs de humano y ratón. En estas regiones de 100kb, tenemos en cuenta todos los picos de ChIP-seq de CTCF presentes en al menos 5 tipos celulares de ratón (Dunham *et al.* 2012) o que sean constitutivos en el caso de humanos (Li *et al.* 2013). Mediante el uso del software Clover (con un valor umbral de “10”) , buscamos dentro de estos picos motivos de CTCF y su orientación. La región de 100kb analizada de cada borde topológico se divide en 20 ventanas de 5 kb, y se representa, para cada orientación, el número de motivos de CTCF que se han encontrado en cada ventana de 5kb en el total de los bordes analizados (Fig. RE_16, A y B). Como control de la distribución de sitios de unión a CTCF, analizamos 1000 ventanas de 100kb centradas en promotores, que se conoce que son regiones enriquecidas también en sitios de unión de CTCF, y realizamos el mismo análisis que para los bordes

topológicos (Fig. RE_16, C y D). En el caso de los bordes topológicos, existe un sesgo en la distribución de los sitios de unión de CTCF a un lado y al otro de dichos bordes en función de su orientación. Así encontramos que los sitios de unión a CTCF aguas arriba de los bordes topológicos, tanto en humanos como en ratón, se encuentran con una mayor frecuencia orientados con la cabeza aguas arriba, mientras que aguas abajo del borde la orientación mayoritaria es la opuesta (Fig. RE_15, A y B). En el caso de los promotores, por el contrario, no encontramos ningún tipo de sesgo en la distribución de los sitios de unión a CTCF en



26 Fig RE_16

Fig. RE_16: Los bordes topológicos en humano y ratón están enriquecidos en la presencia de sitios de unión de CTCF orientados de manera divergente. A y B) El gráfico superior muestra la orientación y número de sitios de unión a CTCF a lo largo de 100 kb alrededor de los bordes topológicos de humano (A) y ratón (B). El gráfico inferior muestra la distribución y media de la localización de los sitios de unión a CTCF en función de su orientación. C y D) El mismo análisis anterior pero centrado las 100 kb en 1000 promotores enriquecidos en CTCF seleccionados aleatoriamente en humano (C) y en ratón (D).

función de su orientación (Fig. RE_15, C y D). Además, el 72% de los bordes en ratón y el 48% de los bordes en humanos, contienen al menos un par de sitios de unión a CTCF orientados de manera divergente.

Estos datos sugieren que la presencia de sitios de unión orientados de manera divergente podría ser una característica específica y general de los bordes de los dominios de asociación topológica presentes en los genomas de mamíferos en este caso.

Discusión

1. DE LA GENÓMICA COMPARATIVA A LA ARQUITECTÓMICA COMPARATIVA.

Durante la historia evolutiva de los seres vivos éstos se han diversificado en una inmensa variedad de formas. La diversidad morfológica en los Metazoos representa el ejemplo más paradigmático y ampliamente analizado. Esta diversidad se ha estudiado durante siglos desde un punto de vista comparativo, atendiendo a aspectos morfológicos, fisiológicos y embriológicos. A partir de la segunda mitad del siglo XX, con el desarrollo de la biología molecular, la biología evolutiva se concentró también en comprobar los aspectos genéticos, y hoy en día en la comparación de genomas completos, en lo que se conoce como “genómica comparativa”.

Hace ya mucho tiempo que sabemos que el material responsable de la herencia de los caracteres y propiedades específicas de los organismos celulares radica en el ADN (Hersey y Chase 1952). Hoy en día conocemos el conjunto del ADN de un organismo como genoma. Esta entidad molecular contiene toda la información necesaria para la generación, funcionamiento y reproducción de los distintos organismo vivos. Esto apoya que la comunidad científica haya puesto enormes esfuerzos en entender y descifrar el código y la lógica de este lenguaje molecular que, en definitiva, alberga los secretos de la vida.

A raíz del Proyecto Genoma Humano (Lander y cols 2001, Gregory y cols 2006), se desarrolló la tecnología capaz de secuenciar genomas completos, y los métodos para el análisis de tan ingente cantidad de información. Este desarrollo tecnológico ha permitido que hoy en día conozcamos el genoma completo de otras especies animales y que cada vez sea más asequible conocer la secuencia genómica completa de multitud de organismos complejos (Holt y col 2002, Aparicio y cols 2002, Waterson y cols 2002, Howe y cols 2013).

La genómica comparativa ha sido y es una herramienta fundamental para progresar en el entendimiento del lenguaje genómico. Ha permitido comprobar como el contenido génico “*per sé*” no es capaz de explicar las diferencias morfológicas ni los distintos niveles de complejidad que existen en el mundo animal (Erwin 2009, Davidson y Erwin 2010). De hecho, la genómica comparativa ha revelado que el contenido génico sobre todo en el caso de los genes de desarrollo está, en gran medida, conservado en todos los animales. Si la

genómica comparativa ha sido importante para entender el “*genoma codificante*”, se ha convertido en una herramienta mucho más importante y esencial para la exploración, caracterización y comprensión del “*genoma no codificante*” (Martínez-Morales 2015, Hardison y Taylor 2012, Beaster-Jones 2012, Meireles-Filho y Stark 2009, Venkatesh y Yap 2005).

Ha sido a través de la comparación de las secuencias genómicas de distintas especies, como por primera vez se han identificado, a escala global, fragmentos de ADN funcionalmente relevantes más allá de las secuencias codificantes y su aparato regulador proximal. Estas comparaciones genómicas han puesto de relieve que el “*genoma no codificante*” se encuentra poblado de elementos funcionales, algunos conservados a lo largo de grandes distancias filogenéticas. La caracterización de estas regiones ha revelado que muchas tienen la capacidad de modular la expresión génica. Debido a la presencia frecuente de estos elementos reguladores a lo largo del “*genoma no codificante*”, hoy en día hablamos de “*genoma regulador*”, definición funcional que hace referencia al conjunto de elementos de ADN del genoma que tienen la capacidad, de una forma u otra, de modular la expresión de los genes.

Si bien la genómica comparativa ha sido la base del estudio del “*genoma regulador*”, esta estrategia está limitada por definición a la hora de dilucidar el aparato regulador completo de un genoma. En su máxima expresión, esta aproximación sólo nos revelaría los posibles elementos funcionales conservados o comunes entre las especies comparadas, aspecto de una increíble relevancia biológica y evolutiva, pero que dejaría ocultos todos los elementos específicos de una especie en particular, aspecto también de gran importancia, pues en ellos radicaría, por ejemplo, la clave de muchas innovaciones morfológicas y funcionales de esa especie. Pero incluso a la hora de encontrar los elementos funcionales conservados entre dos especies, la comparación de genomas tiene grandes limitaciones. A lo largo de la historia evolutiva, los genomas sufren cambios. Entre estos cambios podemos encontrar mutaciones, inserciones, eliminaciones, translocaciones o duplicaciones. Estos cambios pueden ser de pequeña o de gran envergadura, llegando incluso a producirse duplicaciones completas del genoma (Glaseauer y Neuhauss 2014, Cañestro y cols 2013). La cuestión es que los genomas animales, si bien conservan un orden relativo de elementos de ADN en determinados puntos (normalmente complejos génicos), han sufrido una gran cantidad de reorganizaciones. Así pues, no resulta sencillo alinear genomas completos o grandes fragmentos del mismo, y por tanto es factible que un mismo elemento regulador que sí esté presente en dos especies no

sea detectado por encontrarse en una posición relativa diferente en cada especie. Por otro lado, los elementos cis-reguladores son fragmentos de ADN con la habilidad de reclutar determinados factores de transcripción. Hemos mencionado como una misma combinación de factores de transcripción pueda ser reclutada por elementos reguladores con secuencias con bajo porcentaje de identidad. Por lo tanto, un elemento regulador de una especie que tenga su equivalente en otra, incluso estando en la misma posición relativa, puede llegar a no ser detectado por conservación, dado que ambas secuencias pueden ser lo suficientemente diferentes para no alcanzar el umbral de detección.

Estas limitaciones de la genómica comparativa para abordar el estudio del “genoma regulador” se han visto en gran medida complementadas y superadas con la llegada de las técnicas experimentales de detección de sitios de unión de proteínas al ADN, modificaciones epigenéticas y de regiones de cromatina abierta a escala global (ChIP-seq, FAIRE, ATAC-seq, entre otras). De una u otra forma, contando siempre con determinados sesgos, estas técnicas ayudan a localizar potenciales elementos reguladores independientemente de su secuencia, posición o grado de conservación de los mismos. Así pues, estas técnicas revelarán tanto el potencial aparato regulador conservado como específico del genoma de cada especie. Además ayudarán a resolver las limitaciones en la detección de elementos conservados de la comparación genómica. Estas nuevas herramientas revelarían inmediatamente los elementos conservados situados en posiciones relativas equivalentes a pesar de poseer secuencias dispares. Por otro lado, ayudarían a localizar posibles elementos conservados pero que hayan sufrido una reorganización espacial mediante la búsqueda de un perfil equivalente de modificaciones epigenómicas o de unión de factores de transcripción.

Los métodos experimentales van mucho más allá de la simple complementación de la genómica comparativa. El carácter dinámico de las marcas que detectan estas técnicas (modificaciones epigenómicas, unión de factores de transcripción, accesibilidad de la cromatina, etc.) hace que revelen no sólo si se trata o no de un potencial elemento regulador, sino de cuándo y en qué circunstancias este elemento está activo o silenciado. Por lo que, si seguimos hablando de conservación, estas técnicas permiten revelar no sólo la conservación de la presencia de un elemento sino también si este elemento conserva el patrón de activación.

Muchos de estos elementos reguladores son capaces de ejercer su acción reguladora sobre genes situados a grandes distancias. Esto hace que la región de la cromatina con potencial de

albergar elementos reguladores de un gen particular pueda llegar tener una gran extensión. Siguiendo este principio, un solo gen puede estar bajo el rango de acción de varios elementos reguladores y un elemento regulador tiene el potencial de poder actuar sobre varios genes (Montavon y cols 2011, Tena y cols 2011). La acción de los elementos reguladores sobre los promotores se produce mediante interacciones físicas. Esto hace necesario que los elementos cis-reguladores y los promotores diana deban estar localizados con una cierta proximidad física, de forma que cuando sean activados tengan una probabilidad efectiva de interactuar funcionalmente. Esta co-localización entre elementos de ADN distales, se produce porque adquiere una configuración tridimensional particular. De esta forma elementos de ADN lejanos pueden compartir un mismo territorio tridimensional y establecer interacciones funcionales entre ellos con una alta frecuencia, cuando el contexto regulador es el adecuado.

Los genes de desarrollo son proclives a encontrarse localizados en lo que se conoce como “desiertos génicos” (Venter y cols 2001), amplias regiones del genoma con baja densidad de secuencias codificantes de proteínas. Estos desiertos génicos asociados a genes de desarrollo suelen contener multitud de elementos cis-reguladores de la transcripción (Nobrega y cols 2003, Ovcharenko y cols 2005, de la Calle-Mustienes y cols 2005). Los genes de desarrollo generalmente poseen patrones de expresión complejos durante el desarrollo de un embrión lo que se ha asociado a la acción de una gran cantidad de elementos reguladores (de Laat y Duboule 2013, Symmons y Spitz 2013). Esto hace que dichos genes suelen tener asociados paisajes reguladores amplios en extensión y complejos en elementos, que necesitan de una organización espacial robusta y elaborada que dé soporte a sus necesidades transcripcionales. Al igual que la eliminación o incorporación de un módulo regulador al entorno de un gen de desarrollo puede tener importantes consecuencias en el plan corporal final de un organismo, la desorganización espacial de su paisaje regulador podría tener consecuencias drásticas para el proceso de embriogénesis. Es por esto que la organización espacial de la cromatina en relación al paisaje regulador de un gen de desarrollo es crítica.

El plan corporal de un organismo animal depende de la forma en que sus genes de desarrollo son expresados durante el desarrollo embrionario. Cuanto más cercanas filogenéticamente se encuentran dos especies, más similares son sus planes corporales. Esto se debe a que los genes de desarrollo de ambas especies, se comportan (transcripcionalmente hablando) de una forma parecida, debido a que el contenido de sus paisajes reguladores es similar. Si la organización tridimensional de la cromatina es importante para la correcta acción de dichos

paisajes reguladores, esta tendría que mostrar también un determinado grado de conservación y lo mismo ocurriría con los mecanismos moleculares responsables de la misma.

A través del estudio de la organización espacial de los complejos de genes de desarrollo *Six*, desde una perspectiva evolutiva, hemos pretendido contribuir al entendimiento de la importancia de la configuración tridimensional de la cromatina en la organización del paisaje regulador de los genes de desarrollo, su impacto en el control transcripcional de dichos genes, su grado de conservación, y los mecanismos moleculares que pudieran dar lugar a dicha configuración. Esto podría considerarse un ejercicio de “arquitectómica comparativa” o “interactómica comparativa”. Así como la genómica comparativa revela secuencias de ADN conservada entre especies, poniendo de relieve potenciales elementos funcionales esenciales que han debido conservarse; la “arquitectómica comparativa” es un ejercicio que pretende revelar principios organizativos conservados de la cromatina, poniendo de relieve posibles elementos estructurales esenciales para la organización de la información reguladora del genoma.

2. ORGANIZACIÓN 3D DE LOS COMPLEJOS SIX: PEZ CEBRA

Los genes *Six*, como muchos genes de desarrollo, llevan a cabo múltiples funciones en momentos y tejidos distintos durante el desarrollo embrionario de los animales (Kawakami y cols. 2000; Kumar 2009). En consecuencia, estos genes experimentan un comportamiento transcripcional complejo. Esta complejidad transcripcional se asocia a la acción de múltiples elementos cis-reguladores distribuidos a lo largo de la cromatina circundante a los genes *Six* (Splinter y de Laat 2011). En Deuterostomos, los genes *Six* se encuentran organizados en complejos génicos altamente conservados. Esta conservación de la sintenia u orden relativo de los genes, se relaciona con una alta restricción reguladora (Irimia y cols 2012), es decir, que comparten elementos cis-reguladores esenciales que impide la rotura del complejo. La cercanía en la secuencia genómica de los genes *Six* que componen un mismo complejo, los hace susceptibles de compartir una gran proporción de sus paisajes reguladores y por tanto de sus dominios de expresión. Sin embargo, genes los *Six* presente en un mismo complejo desarrollan patrones de expresión muy diferentes durante el desarrollo embrionario, sugiriendo que en cada complejo *Six* existen dos entornos reguladores bien diferenciados.

En el organismo modelo vertebrado pez cebra (*Danio rerio*), encontramos al menos cuatro complejos de genes *Six*: *six3a-six2a*, *six3b-six2b*, *six6a-six1a-six4a* y *six6b-six1b-six4b*. En cada uno de estos complejos encontramos dos patrones de expresión diferenciados (Fig. RE_5). Uno correspondiente a los genes *six3a*, *six3b*, *six6a* y *six6b* que estaría caracterizado principalmente por la expresión en estructuras del cerebro anterior y la retina ventral. El otro patrón de expresión correspondería a los genes *six2a*, *six2b*, *six1a*, *six4a*, *six1b* y *six4b* que se caracterizaría por contener territorios de la vesícula ótica, arcos branquiales, músculos y otros derivados mesodérmicos. Si bien esta descripción de los patrones de expresión se trata de una generalización no exenta de errores e inexactitudes, y a pesar de no haber estudiado los patrones con un detalle celular exhaustivo, sólo trata de poner de manifiesto que efectivamente cada complejo *Six* se encuentra asociado a dos entornos reguladores bien diferenciados, que proporcionan dos patrones de expresión poco o nada solapantes y que de una forma u otra son análogos entre unos complejos y otros.

A través de la aplicación de la técnica de “Captura Circular de la Conformación de la Cromatina”, o 4C, aplicada a los promotores de los genes *Six* y otros elementos circundantes a dichos genes, hemos explorado la organización espacial de la cromatina en estos complejos génicos y su relación con el control transcripcional de los mismos. En el caso del complejo *six2a-six3a*, cuyos promotores se separan 30kb, vemos como los patrones de interacción de cada gen en muestras de embriones a 24hpf abarcan en torno a 500kb en direcciones opuestas (Fig. RE_1 B). Sin embargo, se produce una baja frecuencia de contacto entre ambos promotores. Estos dominios de contacto bien diferenciados correlacionan con las grandes diferencias en el patrón de expresión de *six2a* y *six3a* a 24 hpf. Para cuantificar la distribución de la señal de 4C a lo largo del locus, inferimos un hipotético borde topológico (BTP) y calculamos el porcentaje de señal a uno y otro lado de dicho borde. La mayor parte de la señal de *six3a* (86%) queda contenida en la región aguas arriba del BTP, mientras que la mayor parte de los contactos de *six2a* (64%) corresponde a regiones situadas aguas abajo. Esto sugiere una partición de locus en dos Dominios Topológicamente Asociados o TADs.

En consonancia con la existencia de dos Dominios Topológicos, cuando buscamos aguas arriba del hipotético borde topológico, encontramos elementos cis-reguladores con actividad potenciadora que recapitulan parcialmente el patrón de expresión de *six3a*. Por otro lado, aguas abajo de dicho borde encontramos elementos potenciadores con actividad transcripcional concordante con el patrón de *six2a* (Fig. RE_2 C). Además, cuando analizamos el patrón de contacto de algunos de estos elementos (concretamente el III y el

VI), observamos como son capaces de interactuar fuertemente con el promotor del gen situado en su mismo Dominio Topológico, mientras que los contactos en el dominio contiguo, más allá del borde topológico, son residuales (Fig. RE_2 A). Por otro lado, los elementos reguladores no sólo contactan con su promotor diana, sino que comparten con él multitud de interacciones. Esto sugiere que los contactos potenciador-promotor probablemente no se producen de forma individual, sino que ambos elementos forman parte de una compleja estructura donde se dan multitud de interacciones entre una variedad de elementos cis-reguladores y promotores como ocurre en otros loci (Montavon y Duboule 2012; Hughes y cols 2014). Por tanto, la partición del complejo en dos Dominios Topológicos ayudaría a generar dos paisajes reguladores independientes, evitando en gran medida el intercambio de información reguladora, y facilitando el establecimiento de los contactos correctos entre promotores y potenciadores (Bouwman y de Laat 2015).

El limitado número de estudios sobre la dinámica de los TADs a lo largo del proceso de diferenciación celular está circunscrito a mamíferos, y se observa poca variación en los mismos durante dicho proceso (Nora y cols 2012; Dixon y cols 2012). En concordancia con esta apreciación, cuando estudiamos el patrón de interacción de los genes *six2a-six3a* en distintos momentos del desarrollo del pez cebra (incluyendo aquellos donde ambos genes no se expresan, como “Dome” o 64 células), observamos como el principio organizativo del complejo *six2a-six3a* en dos TADs diferenciados se mantiene constante (Fig. RE_3). Aunque no podemos descartar que se den fluctuaciones en interacciones concretas asociadas a la dinámica de activación de elementos reguladores que puedan escapar a los límites de resolución de la técnica de 4C aplicada sobre embriones completos, no esperaríamos que estas fluctuaciones funcionales cambiaran de manera drástica el principio organizativo bimodal del complejo *six2a-six3a*. Los datos sugieren que la arquitectura tridimensional del complejo *six2a-six3a* está presente incluso antes de que exista ninguna transcripción. De hecho el 4C de *six3a* en embriones de 64 células, a pesar de su poca calidad y la necesidad de réplicas biológicas, sugiere que los contactos distales de *six3a* podrían estar presentes desde el inicio de la embriogénesis. Este pre-ensamblaje del aparato regulador, facilitaría la activación precisa de los genes *Six* una vez el contexto regulador fuese el adecuado aportando robustez al sistema. Este hecho ha sido puesto de manifiesto en otros loci de mamíferos (Jin y cols 2013; Noordermeer y cols 2014) y *Drosophila* (Ghavi-Helm y cols 2014).

Al igual que el complejo *six2a-six3a*, el resto de complejos *Six* presente en el pez cebra parecen estar divididos en dos entornos reguladores diferenciados, como sugiere el patrón de expresión de los genes presentes en un mismo complejo (Fig. RE_8 A-D). El estudio en embriones de 24hpf de los contactos establecidos por cada gen del resto de complejos *Six* sugiere que, al igual que el complejo *six2a-six3a*, el resto de complejos se organiza en dos entornos físicos diferenciados separados por un borde topológico. Así un dominio topológico corresponde a los genes *six3/6* (*six3b*, *six6a* y *six6b*) y el resto a los genes *six1/2* (*six2b*, *six1a* y *six1b*) y los *six4* (*six4a* y *six4b*). Esta división topológica coincide con las diferencias y similitudes transcripcionales. Así como los genes *Six* contenidos en dominios topológicos opuestos (p.e: *six6a* y *six1a*) desarrollan patrones de expresión muy diferentes, los pertenecientes a un mismo dominio (p.e: *six1a* y *six4a*) comparten gran parte de sus dominios de expresión. Esto se refleja en la gran similitud que existe entre los patrones de contactos de los genes *Six* de un mismo dominio topológico, en comparación con genes *Six* que se encuentran en dominios topológicos distintos, incluso cuando no hay diferencias sustanciales en la distancia lineal entre unos y otros (Fig. RE_9, D). Cuando analizamos la estructura de la cromatina de estos complejos *Six* en distintos estadios del desarrollo, la arquitectura bimodal permanece constante y se encuentra presente también incluso antes de que tenga lugar la activación de los genes.

Todos estos datos sugieren que en los complejos *Six* de pez cebra encontramos un principio organizador común que consiste en la generación de dos dominios físicos independientes separados por un borde o barrera entre los genes *six3/6* y *six1/2*. Esta división física o topológica de la cromatina en este punto, favorecería la implementación de dos entornos reguladores separados en gran medida que daría lugar a la expresión diferencial que encontramos en los miembros de un mismo complejo.

3. ORGANIZACIÓN 3D DE LOS COMPLEJOS SIX: UNA MIRADA EVOLUTIVA

La organización de los genes *Six* en complejos génicos (*Six3/6-Six1/2-Six4/5*) aparece en muchos filos animales evolutivamente distantes (Irimia y cols 2012). Como consecuencia de una duplicación ancestral del genoma completo, la mayoría de los vertebrados contienen dos

complejos *Six* (*Six2-Six3* y *Six6-Six1-Six4*). En la base del linaje de los teleósteos tuvo lugar una duplicación genómica extra (Taylor y cols 2001), razón por la cual en el pez cebra encontramos los cuatro complejos estudiados. El hecho de que los cuatro complejos de pez cebra compartan el mismo principio organizativo tridimensional, sugiere que este pudo estar presente en algún momento en el complejo ancestral. Para explorar esta hipótesis decidimos caracterizar la arquitectura tridimensional de los dos complejos *Six* presentes en ratón (*Mus musculus*) y del único complejo presente en el erizo de mar (*Strongylocentrotus purpuratus*).

Tomando como puntos fijos los promotores de los genes *Six* de ratón (*Six2*, *Six3* y *Six6*, *Six1* y *Six4*) realizamos 4C en muestras de embriones completos en estadio 14.5 dpf para todos ellos además de en estadio 9.5 para los genes *Six2* y *Six3*. Los patrones de contacto revelan de nuevo una partición organizativa de ambos complejos génicos en dos dominios topológicos, con el borde localizado entre los genes *Six3/6* y *Six1/2* (Fig. RE_10 A y B). En el caso del complejo *Six2-Six3* este principio organizativo se mantiene en el estadio 9.5 dpf, con cambios moderados en interacciones puntuales (Fig. X). Esta partición topológica correlaciona de nuevo con el patrón de expresión divergente observado en los genes situados a uno y otro del borde topológico. Así, los genes *Six1/2* tienen un patrón de expresión muy distinto a los genes *Six3/6* pero muy parecido al gen *Six4* con el compartimento dominio topológico. La disponibilidad de datos de HiC en células madre embrionarias (ES) de ratón nos permite comprobar si, en efecto los complejos *Six* de ratón se encuentran divididos en dos Dominios Topológicamente Asociados (TADs) con el borde situado entre los genes *Six3/6* y *Six1/2*. Podemos comprobar como efectivamente ambos complejos se sitúan en la frontera entre dos TADs, y como el borde predicho a partir de los datos de 4C coincide con la localización del borde entre dichos TADs. Además observamos como la tendencia y extensión de los contactos de cada promotor obtenidos mediante 4C se ajustan, en gran medida, a la extensión y límites marcados por los TADs. El contexto regulador de las células ES puede considerarse muy diferente al que podemos encontrar de forma mayoritaria en embriones del estadio 14.5 dpf, aun así, los aspectos estructurales básicos de la organización tridimensional de los complejos *Six* parecen independientes de dicho estado regulatorio.

El estudio en ratón sugiere que, al menos, parte del linaje de los vertebrados puede compartir los principios organizativos tridimensionales alrededor de los complejos *Six*. Dado que este principio organizativo estaría presente en ambos complejos *Six* de ratón, es posible que esta arquitectura existiese antes de la duplicación genómica que tuvo lugar en la base de los

vertebrados. Para verificar esta posibilidad estudiamos, mediante 4C, la arquitectura del único complejo *Six* presente en el erizo de mar, *Strongylocentrotus purpuratus*.

El erizo de mar es un equinodermo, y por lo tanto se localiza en una de las ramas de los deuterostomos más alejadas filogenéticamente de los vertebrados. *S. purpuratus* contiene un único complejo *Six*. Durante la fase larvaria tan sólo los genes *Sp-Six3/6* y *Sp-Six1/2* son expresados en algún momento, mientras que del gen *Sp-Six4/5* no se detectan niveles de transcrito significativos. El gen *Sp-Six3/6*, análogamente a vertebrados (Steinmentz y cols 2010), se expresa durante la gastrulación tardía (48hpf) en territorios apicales neurogénicos donde se desarrollaran más tarde células sensoriales. Por otro lado, y de una forma muy diferente, el gen *Sp-Six1/2* se expresa en territorios mesodérmicos del arquenteron y células esqueletogénicas. Los 4C de los genes *Six* de *S.purpuratus* en muestras de embriones completos a 48hpf revelan cómo *Sp-Six3/6* y *Sp-Six1/2* apenas comparten dominios de contacto (Fig. RE_10 C). Sin embargo, dada la gran distancia lineal que existe entre ambos genes (>200kb) no podemos asegurar que este contacto diferencial sea debido a la formación de dos dominios topológicos diferenciados y no de un simple efecto de la distancia entre ambos promotores. Si bien el hecho de encontrar contactos significativos de *Sp-Six1/2* más allá de dominio de contacto de *Sp-Six3/6*, sugiere que para evitar el solapamiento mayoritario entre ambos paisajes reguladores es necesario algo más que la simple distancia lineal entre los promotores. El uso de un mayor número de puntos fijos sobre los que realizar 4C a lo largo del complejo *Six* de *S.purpuratus* sería fundamental para aclarar esta cuestión.

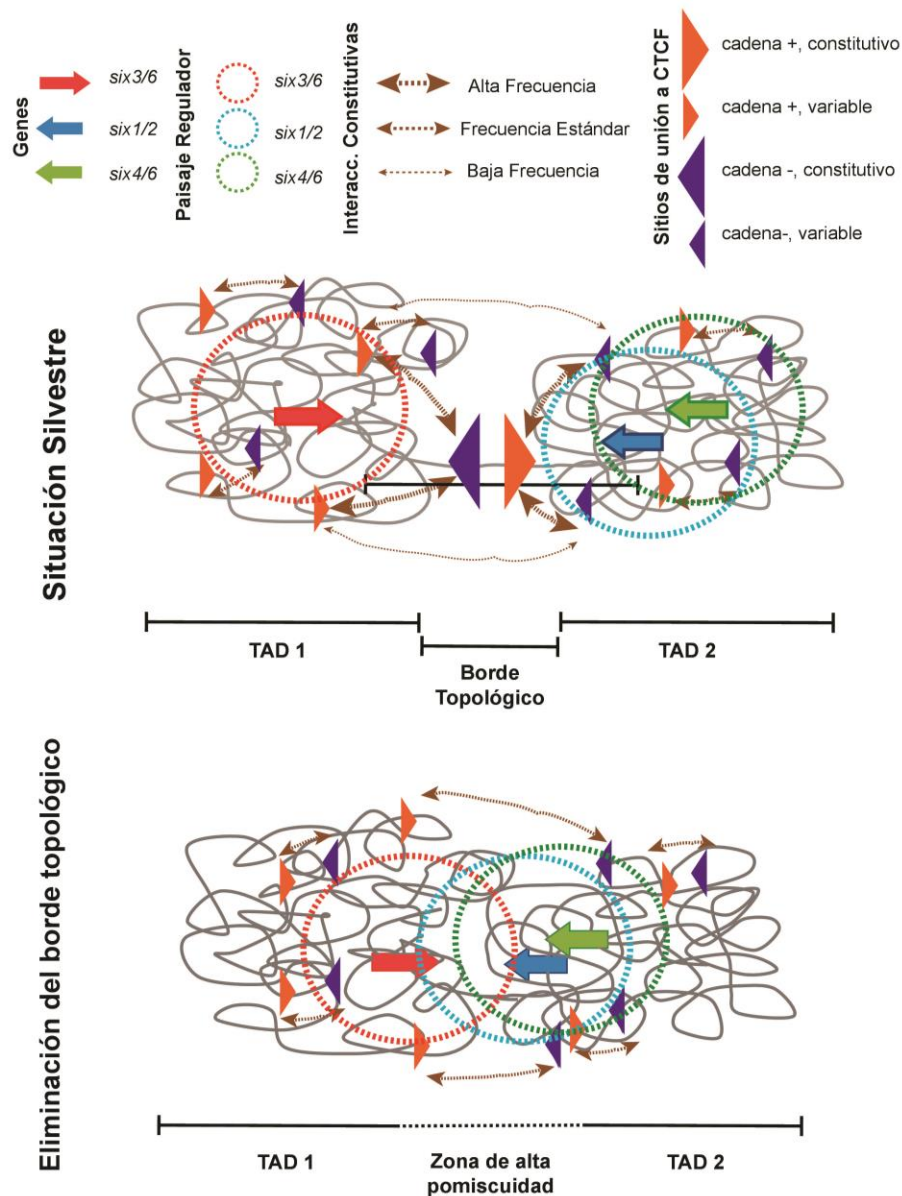
Un aspecto fundamental del principio organizativo que observamos en los complejos *Six*, es que esta arquitectura bimodal parece ser independiente del contexto regulador. Así en pez cebra encontramos esta organización en momentos muy distintos del desarrollo donde los genes *Six* muestran un comportamiento transcripcional muy diferente. Lo mismo ocurre en el caso de ratón donde encontramos esta organización mediante 4C en al menos dos momentos distintos del desarrollo (9.5 y 14.5 dpf), y basándonos en los datos de HiC disponibles, está también presente en células madre embrionarias y tejido cerebral adulto. Además si comprobamos los datos de HiC publicados tanto en células madre embrionarias como en fibroblastos humanos (Dixon y cols 2012), la organización 3D de los complejos *Six* en dos TADs contiguos también parece estar presente y ser independiente del contexto regulador en humanos. En consecuencia, todos estos datos sugieren que las fuerzas moleculares responsables en última instancia de la organización 3D básica de los complejos *Six*

probablemente son fuerzas estructurales ejercidas por elementos moleculares (cis y trans) constitutivos, que no dependen de un programa regulador concreto. Esto descarta, en gran medida, a las interacciones funcionales mediadas por factores de transcripción específicos de tejido (entre elementos potenciadores y promotores diana) como responsables principales de la organización 3D de los complejos *Six*, al menos en las especies analizadas. Sin embargo, estas interacciones funcionales, podrían contribuir a reforzar o relajar la organización de la cromatina alrededor de los complejos *Six* para favorecer o penalizar la actividad transcripcional de los genes del complejo. Por otro lado, como hemos mencionado, los datos sugieren que la organización 3D de los complejos *Six* se sustenta en interacciones estructurales. Estas interacciones estructurales podrían estar mediadas por proteínas arquitectónicas constitutivas. Por tanto proteínas tales como CTCF o cohesina (Gómez-Díaz y Corces 2014) serían buenos candidatos a ser en gran medida los elementos *trans* involucrados en mantener el principio organizativo de los complejos *Six*.

4. LA FUNCION DEL BORDE TOPOLÓGICO

La organización conservada de los complejos *Six* en dos TADs a lo largo de la evolución, nos ha llevado a tratar de dilucidar cuál es el impacto que dicha organización tridimensional tiene en la regulación de los genes *Six* durante el desarrollo embrionario. La base de esta organización tridimensional radica en la generación de dos entornos físicos aislados a partir de un borde topológico situado entre los genes *Six3/6* y *Six1/2*.

Como ejemplo para la evaluación del papel funcional del borde topológico nos centramos en el complejo *six3a-six2a* de pez cebra, el más ampliamente caracterizado desde un punto de vista estructural y regulador. Generamos líneas transgénicas con el BAC DK74B2 (que abarca la región que comprende ambos genes y parte de sus paisajes reguladores), donde sustituimos ambas secuencias codificantes por genes reporteros (*six3a* por *mCherry*, y *six2a* por *GFP*). Además generamos otra línea transgénica con una versión modificada del BAC DK74B2, donde eliminamos 18kb alrededor del borde topológico. Basándonos en los datos ChIP-seq generados en nuestro laboratorio, esta región eliminada no contiene ninguna señal exigentica significativa de H3K27ac en la primeras 48 hpf. Este dato sugiere que no existen elementos reguladores relevantes a lo largo de las 18kb eliminadas, si bien no podemos descartarlo con total seguridad.



27 Fig D_1

Fig D_1: Modelo propuesto para la organización tridimensional de los complejos *Six*. Los genes *Six3/6* por un lado y los genes *Six1/2* y *Six4/5* pertenecientes a un mismo complejo estarían organizados de dos TADs diferenciados y separados por un borde topológico caracterizado por la presencia de sitios de unión constitutivos y estables a CTCF orientados de manera divergente. Cada TADs se sustentaría en cierta medida por la acción de interacciones intra-TAD constitutivas, a las que también contribuiría CTCF. Lo sitios divergentes de CTCF en el borde topológico contribuiría a la definición territorial de ambos TADs, penalizando los contactos entre sitios de unión a CTCF presentes en TADs opuestos e impidiendo la formación de interacciones constitutivas entre ambos TADs. Esta separación en dos TADs de los complejos *Six* genes *Six3/6* con un paisaje regulador bien diferenciado de aquel asociado a los genes *Six1/2* y *Six4/5*. La ausencia del borde topológico, si bien no produciría necesariamente una fusión de ambos TADs, ya que todavía estarían sustentados por las interacciones intra-TADs, pero sí eliminaría en gran medida las restricciones físicas presentes en la frontera entre ambos TADs, generando una región sin constricciones donde todos los elementos tienen el potencial de interactuar entre ellos. Así la eliminación del borde topológico en los complejos *Six* podría producir un solapamiento parcial de los paisajes reguladores de los genes de un mismo complejo con importantes consecuencias en su comportamiento transcripcional, y en consecuencia en el desarrollo embrionario del organismo.

La comparación de ambas líneas transgénicas a 5 dpf revela que en ausencia del borde topológico, *six3a* reduce enormemente su expresión a la vez que *six2a* es capaz de reemplazar, en términos de expresión, a *six3a* en algunos de los territorios perdidos por este último. Teniendo en cuenta las limitaciones que esta aproximación experimental tiene respecto de la modificación del locus endógeno, estos datos sugieren que *six2a* es capaz de competir por parte del aparato cis-regulador de *six3a*, y que en ausencia del borde topológico *six2a* puede secuestrar parcialmente información reguladora de *six3a*. Por tanto el borde topológico resultaría crítico para evitar la competencia de los promotores por elementos reguladores situados al otro lado del borde.

La gran ventaja de la aproximación con BACs respecto de las modificaciones endógenas es que los efectos observados en el patrón de expresión son debidos necesariamente a cambios en cis (bien por ausencias de elementos reguladores específicos o bien por cambios estructurales), ya que en realidad no estamos modificando la expresión de ningún factor de transcripción. Si nuestra hipótesis sobre la competencia de los promotores *six2a* y *six3a* por elementos reguladores en ausencia del borde topológico es cierta, esperaríamos encontrar cambios en la configuración tridimensional del complejo y por tanto en el patrón de contactos de los elementos reguladores.

Esto es lo que observamos cuando realizamos 4C sobre los elementos reguladores III y VI del complejo *six2a-six3a*. En líneas transgénicas transitorias con los BACs linealizados, observamos que en la versión completa del BAC, ambos elementos respetan el límite marcado por el borde topológico (como hicieran los elementos endógenos). Sin embargo, en la versión modificada del BAC, observamos que en ausencia de borde topológico los elementos III y VI son capaces de contactar ambos promotores con una frecuencia similar. Este dato apunta a que, subyacente a los cambios de expresión observados, encontramos un cambio en el principio organizativo fundamental del complejo. En ausencia del borde topológico, no existiría una separación o constricción física eficaz de los entornos reguladores de *six2a* y *six3a* favoreciendo la interferencia y el intercambio de elementos reguladores, al menos de los situados cercanos al borde (Fig. D_1).

Existe un limitado número de ejemplos en la literatura donde se haya estudiado el efecto de la eliminación del borde de un TAD en la organización global de la cromatina y su impacto sobre la regulación de los genes circundantes (Nora y cols 2012, Lupiáñez y cols 2015). Aunque muy limitado, este estudio del papel funcional del borde topológico presente en el

complejo *six2a-six3a*, representa una de las pocas aproximaciones del impacto funcional de los bordes topológicos en el contexto del desarrollo embrionario de un organismo.

Por otro lado, el trabajo con BACs ha arrojado información sobre otro aspecto distinto de la organización 3D de los paisajes reguladores. Las líneas estables recientemente obtenidas de la versión completa del BAC DKEY-74B2, nos permitió obtener perfiles de contacto de 4C desde los promotores de *six2a* y *six3a* contenidos en el BAC exclusivamente (los cebadores DpnII se diseñaron en la región codificante del gen de la *GFP* y la *mCherry* respectivamente). Estos perfiles reproducen cualitativamente los perfiles de los promotores endógenos en referencia a la región abarcada en el BAC. Lo interesante es que además establecen contactos con el nuevo entorno genómico donde cada BAC se ha insertado (Fig. RE_12). Sin embargo, estos contactos son de una intensidad mucho menor que los establecidos con su entorno natural. Si bien se trata de un dato aislado, este hecho sugiere que la aparición alrededor de un gen de desarrollo de un nuevo entorno genómico no relacionado evolutivamente no supondría necesariamente la integración automática de dicho entorno en el paisaje regulador del gen. Desde el punto de vista evolutivo este hecho tendría algunas implicaciones importantes. Las reorganizaciones genómicas consecuencia de las recombinaciones meióticas, podrían aportar potencialmente grandes cantidades de “nueva cromatina” al paisaje regulador de un gen de desarrollo sin producir necesariamente un impacto regulador automático importante. Sin embargo esta “nueva cromatina”, con el tiempo, podría actuar como sustrato diferencial para la aparición de “novedades reguladoras” (nuevos elementos cis-reguladores) y optimizar contactos productivos con el promotor del gen. En última instancia estos nuevos contactos funcionales producirían nuevos comportamientos transcripcionales del gen de desarrollo y en consecuencia posibles innovaciones morfológicas y funcionales en el organismo.

5. UNA POSIBLE HUELLA MOLECULAR EN LOS BORDES TOPOLÓGICOS

Desde que por primera vez se aplicaran las técnicas de “Captura de la Conformación de la Cromatina” a gran escala (5C y HiC) al genoma de mamíferos (Nora y cols 2012; Dixon y cols 2012) sabemos que el genoma de los mismos se organiza en Dominios Topológicamente Asociados o TADs. Estos TADs, que tienen un tamaño medio de unas 800kb en mamíferos, se caracterizan porque los elementos de ADN contenidos en él, tienden a interactuar entre

ellos con una frecuencia significativamente mayor que con elementos fuera del mismo. Esta organización del genoma en TADs, no parece ser exclusiva de mamíferos, ya que el genoma de *Drosophila* también se encuentra organizado en estructuras similares aunque de menor tamaño (Sexton y cols 2012, Hou y cols 2012). Aunque el análisis a gran escala de la organización tridimensional del genoma se ha aplicado de momento a un reducido número de especies, la organización en TADs no es una propiedad universal de los organismos eucariotas ya que no aparece en *Arabidopsis* (Grob y cols 2014) y parece que tampoco de forma general en el nematodo *C. elegans* (Crane y cols 2015). Si bien la pérdida de estas estructuras en estos dos últimos organismos puede estar asociada a que su proceso de desarrollo embrionario no requiere de elementos cis-reguladores de larga distancia (Dekker J. y Heard E. 2015).

La importancia de los TADs radica en que es posible que se trate de unidades funcionales del genoma (Lonfat y Duboule 2015, Le Dily y Beato 2015, Dekker y Heard 2015). Así en mamíferos y en *Drosophila* los TADs correlacionan con marcas epigenómicas funcionales tales como modificaciones de histonas, con unidades temporales de replicación (Pope y cols 2014), con dominios de asociación a lámina nuclear, etc. Sin embargo, a pesar de su importancia poco se conoce de los mecanismos moleculares responsables de tal organización. Una característica fundamental, que desprende el análisis en mamíferos, es que pese a que los contactos específicos en el interior de un TAD pueden ser dinámicos, los límites o bordes de muchos TADs se mantienen, en gran medida, estables a lo largo de un proceso de diferenciación celular y están altamente conservados entre mamíferos (Phillips-Cremins y cols 2013, Lieberman-Aiden y cols 2009, Dixon y cols 2012). Por lo tanto entender los mecanismos moleculares que subyacen el establecimiento de un borde entre dos TADs resulta capital para comprender como se forman y funcionan estas estructuras organizativas del genoma.

La “arquitectómica comparativa” aplicada a los complejos *Six* de pez cebra, ratón y erizo de mar, nos ha revelado un principio organizativo común cuya característica fundamental es la existencia de un borde topológico dentro del complejo que divide al mismo en dos dominios físicos distintos, ayudando a generar dos paisajes reguladores diferenciados. Este principio organizativo está también presente en humanos, como se desprende de los datos públicos de HiC en células embrionarias humanas. La conservación del borde topológico en el complejo *Six* a esta escala evolutiva, supone un escenario sin precedentes para la búsqueda de determinantes moleculares responsables de la formación de dichos bordes ancestrales.

La estabilidad respecto del contexto regulador del borde topológico de los complejos *Six*, sugiere que dicho borde es producto de elementos moleculares estructurales más que de factores de transcripción tejido específicos. Por esta razón, en la búsqueda de los determinantes del borde nos centramos en la búsqueda sitios de unión a CTCF (factor de unión a sitios CCCTC). La proteína CTCF se encuentra enriquecida en los bordes topológicos y está involucrada en interacciones constitutivas de larga distancia (Phillips-Cremins y cols 2013), interacciones que tienen lugar principalmente entre sitios de unión orientados de forma convergente (Rao y cols 2014). Esta proteína se encuentra además presente en todos los organismos donde se han detectado TADs además de en *S.purpuratus*, y sin embargo parece haberse perdido en *C.elegans* (Heger y cols 2012) y no se ha detectado en *Arabidopsis*, cuyos genomas no están organizados en TADs (Crane y cols 2015). El análisis de los datos de HiC y ChIP-seq de CTCF en distintos tipos celulares de humano y ratón, reveló que en la posición del borde topológico de todos sus complejos *Six* aparecían sitios de unión a CTCF constitutivos (presentes en diversos tipos celulares) que además presentaban una orientación divergente. Entre humanos y ratón estos sitios se encontraban conservados en términos de secuencia. La predicción bioinformática de sitios de unión a CTCF asociada a picos de ATAC-seq en los complejos *Six* de pez cebra reveló la presencia de una huella molecular conservada a la altura del borde topológico: dos sitios de unión a CTCF orientados de forma divergente. Una disposición similar de sitios de unión a CTCF parece estar presente en el borde del complejo *Six* de erizo de mar. Si bien este dato se basa sólo en una predicción bioinformática que no está apoyada por ningún dato experimental. En general los datos generados en este trabajo sugieren que la disposición en tándem de sitios de unión a CTCF orientados de forma divergente es una huella molecular común al borde topológico de los complejos *Six*.

El análisis de esta configuración específica de sitios de unión a CTCF a escala global en los bordes topológicos de humano y ratón, reveló que dichos sitios están preferentemente orientados hacia fuera del borde (con la cabeza orientada al lado contrario del borde). Además un 72% de los bordes en humano y un 48% de los bordes de ratón contienen al menos una pareja de sitios de unión a CTCF orientados de manera divergente. Este dato revela que esta huella molecular puede ser una característica compartida por los bordes topológicos en general. Nuestra hipótesis es que esta orientación específica de los sitios de unión constitutivos de CTCF ayudaría a generar interacciones estructurales en direcciones opuestas, favoreciendo la generación de dos dominios físicos independientes y penalizando

las interacciones en sentido contrario. En la dirección de esta hipótesis surgió recientemente un trabajo (Guo y cols 2015), donde el simple cambio en la orientación de un sitio de unión a un CTCF es capaz de cambiar la orientación general de la topología de la cromatina circundante.

Estamos aún lejos de comprender los principios moleculares, funcionales y evolutivos de los Dominios Topológicamente Asociados. La definición de los TADs, de momento, es una definición descriptiva de carácter físico (regiones de ADN con una mayor frecuencia de interacción entre sí). Eso sí, en un limitado número de estudios se ha mostrado que dichas estructuras correlacionan con muchos aspectos funcionales y evolutivos del genoma. Muchos tratamos de buscar los principios biológicos, moleculares y funcionales que se encuentran detrás de estas estructuras. Si bien, no todos los TADs que encontramos en un genoma tienen que compartir necesariamente dichos principios ni servir al mismo propósito funcional. Un aspecto muy característico de los TADs son sus límites o bordes, por su conservación evolutiva y a lo largo de un proceso de diferenciación y por el enriquecimiento que en ellos se da de determinadas huellas moleculares. La contribución real de un borde topológico a la estructura global de un TAD es un fenómeno poco caracterizado y en general se desconoce. Hasta qué punto un borde topológico es un elemento causal de los TADs o una mera consecuencia de las interacciones intra-TADs es todavía una pregunta sin resolver. Si bien el grado de especialización, congruencia y conservación de la marcas moleculares asociadas específicamente a los bordes topológicos, invitan a pensar que son algo más que una simple consecuencia. Este trabajo sugiere, con todas sus limitaciones, que el borde topológico *per se* presente en los complejos *Six* contribuye a definir con precisión el límite entre dos TADs contiguos, evitando en gran medida contactos promiscuos entre los paisajes reguladores de los distintos genes miembros de un mismo complejo *Six*. Este papel en la “definición del límite” puede parecer un aspecto menor dentro de la estructura global de un TAD pero resulta fundamental cuando en la frontera del mismo encontramos genes de desarrollo cuyo control transcripcional debe ser estricto para un correcto desarrollo embrionario. En cualquier caso el estudio de la contribución estructural y funcional de los bordes topológicos a los TADs requiere aún de un estudio más sistemático y en contextos más “in vivo”. Estos estudios necesitarán de la perturbación y modificación endógena de los genomas para lo que la tecnología CRISPR/Cas9 resultará esencial en el futuro próximo. Por otro lado la aproximación mediante BACs que hemos hecho en este trabajo, si bien muy limitada, tiene otra ventaja para el estudio de la contribución de los bordes topológicos a la organización del

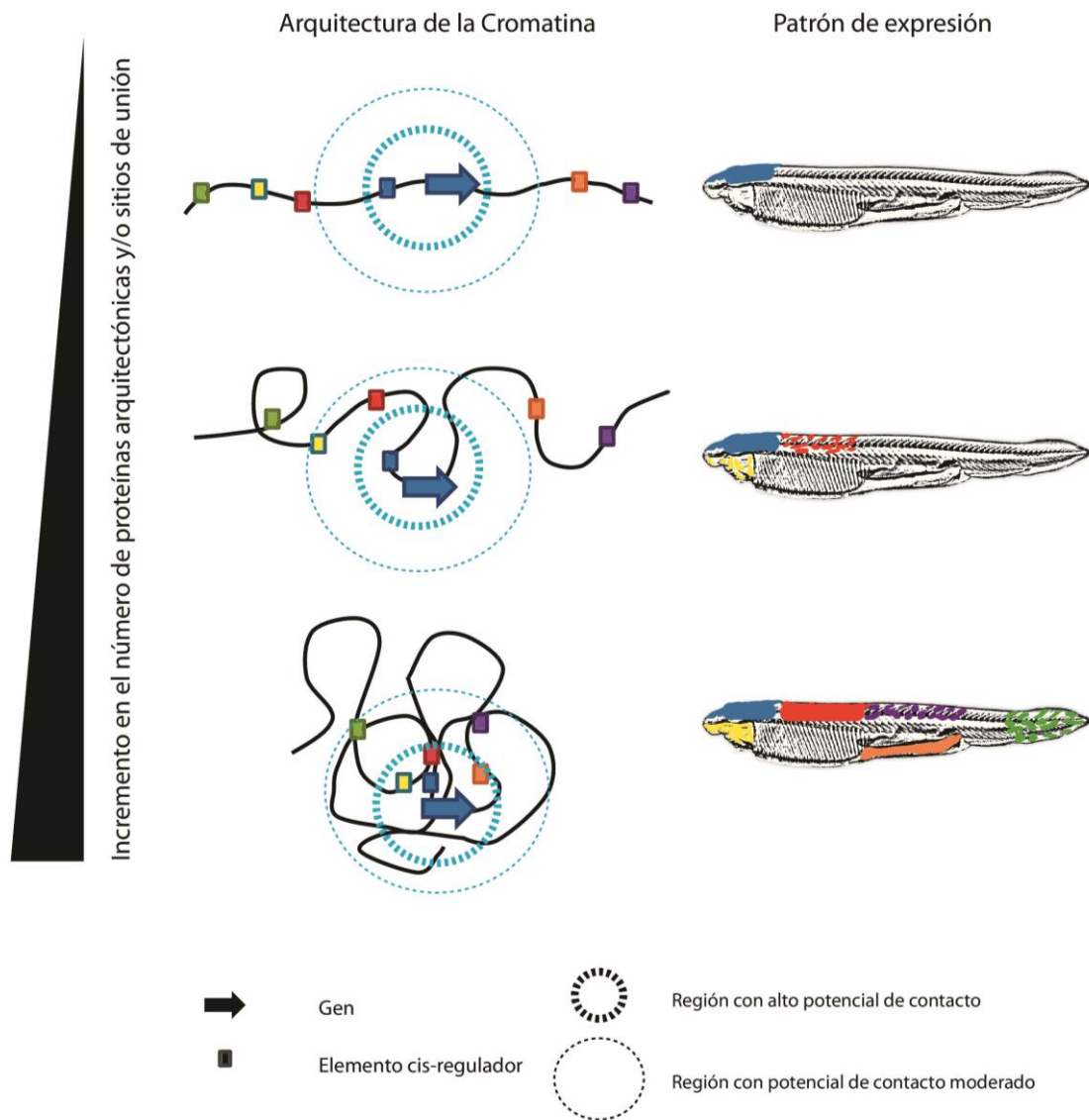
genoma. Esta ventaja consiste en que estamos generando líneas donde introducimos ectópicamente un borde topológico conservado evolutivamente. Esto genera un modelo experimental ideal para contestar a preguntas como ¿Es esto suficiente para generar dos TADs?, ¿En qué manera un borde topológico *per se* modifica la estructura de la cromatina circundante?, ¿Cuáles son las consecuencias transcripcionales sobre los genes vecinos?, etc. Sin duda queda aún mucho por conocer sobre el papel de la organización 3D de la cromatina en el lenguaje y la función de los genomas.

6. INTERACCIONES ESTRUCTURALES, ORGANIZACIÓN 3D DE LA CROMATINA, PLEIOTROPIA Y EVOLUCIÓN.

Uno de los aspectos más apasionantes de la historia evolutiva de los Metazoos es que la inmensa variabilidad morfológica y funcional que encontramos se construye a partir de un limitado y ampliamente conservado grupo de genes de desarrollo. Este hecho es posible porque cada uno de estos genes tiende a realizar una amplia variedad de funciones en distintos momentos y contextos del proceso de desarrollo embrionario de un organismo. Este fenómeno se conoce como pleiotropía. En realidad este término hace referencia a cuando varias características fenotípicas dependen de un solo gen, y se acuñó por primera vez hace más de un siglo (Stearns 2010). El grado de pleiotropía de un gen tendería a restringir la capacidad de cambio de su secuencia codificante (Molodtsova y cols 2014). Un cambio en la secuencia codificante podría afectar a la función en todos los contextos donde el gen de desarrollo tuviera alguna función, generando una perturbación reguladora difícil de asumir por el organismo. Esto dejaría a las modificaciones en las secuencias reguladoras de los genes como principal mecanismo generador de diversidad y complejidad en los Metazoos (Britten y Davidson, 1971; King y Wilson, 1975). La naturaleza modular de los mecanismos cis-reguladores permite que la modificación, pérdida o adquisición de nuevas secuencias reguladoras tenga un impacto más concreto y específico en el comportamiento transcripcional del gen de desarrollo. Esta perturbación específica tiene mayores probabilidades de ser asumida por el programa de desarrollo de un organismo, generando cambios sutiles de la morfología y la función del embrión.

Entendemos como paisaje regulador de un gen a la extensión de cromatina que alberga todos los elementos en *cis* de ADN que de una forma u otra controlan la actividad transcripcional de dicho gen. La capacidad de generar paisajes reguladores complejos determinará la cantidad de información reguladora que pueda ser asociada a un gen en particular, y por tanto su potencial grado de pleiotropía. Esta capacidad dependerá de la habilidad para poner en proximidad física múltiples regiones de cromatina, generando un amplio sustrato genómico que incremente las posibilidades de aparición de elementos *cis*-reguladores con probabilidad de ejercer contactos productivos con el promotor de un gen diana. Por tanto la generación de estructuras 3D complejas y específicas donde intervengan un gran número y extensión de regiones genómicas proporcionaría la base para el incremento en la frecuencia de adquisición de elementos reguladores. Una vez adquirida, si una información reguladora supone una ventaja en la capacidad de supervivencia del individuo, tendería a permanecer, fijando en cierta medida la arquitectura responsable de la asociación entre esa información reguladora y su promotor diana.

En este trabajo y en otros (Dixon y cols 2012, Rao y cols 2014) hemos visto, que al menos una parte importante de la organización 3D de la cromatina no depende del contexto regulador y por tanto, sugiere que los elementos principalmente responsables de la misma son elementos constitutivos. La proteína arquitectónica constitutiva CTCF ha sido uno de los factores más estudiados en el establecimiento de interacciones estructurales de larga distancia. Una hipótesis que explicaría la generación de paisajes reguladores complejos sería una primera fase donde aparecerían distintos sitios de unión a CTCF a lo largo de un locus genómico. Esto generaría una organización tridimensional particular en dicho locus liderada por los contactos entre sitios de unión a CTCF. A priori, no tendría por qué tener ninguna consecuencia funcional pero este tipo de organización 3D pondría en cercanía física múltiples elementos genómicos entre sí, entre los cuales podría estar alguna secuencia codificante. Esta nueva arquitectura constitutiva incrementaría la cantidad de cromatina “accesible” al promotor de dicha secuencia codificante, aumentando la probabilidades de ser influenciado por una mayor cantidad de información reguladora. Si las nuevas relaciones establecidas entre elementos reguladores y gen diana supusieran alguna ventaja para el individuo, esto fijaría y optimizaría la organización constitutiva generada por interacciones entre sitios de unión a CTCF.



28 Fig D_2

Fig. D_2: Relación entre la capacidad para generar relaciones 3D complejas en la cromatina y el potencial para genera diversidad funcional. Un incremento en el número y tipo de proteínas arquitectónicas y/o en sus sitios de unión favorecería la elaboración de organizaciones 3D en la cromatina cada vez más complejas. Físicamente esto se traduciría en que sería posible poner, de manera específica, cada vez más regiones de la cromatina en proximidad espacial relativa generando amplias redes de interacción entre muchas regiones que tendrían potencial para establecer contactos funcionales. Este fenómeno permitiría la elaboración de paisajes reguladores más grandes, diversos y complejos que podrían asociarse a “genes de desarrollo”, favoreciendo la generación de redes de control génico durante el desarrollo a su vez más grandes, diversas y complejas lo que redundaría en el potencial de un organismo para innovar morfológica y funcionalmente.

Siguiendo esta hipótesis, el incremento en el número de proteínas arquitectónicas y/o sitios de unión de las mismas produciría un aumento en la capacidad de generar estructuras 3D de la cromatina cada vez más complejas, permitiendo la asociación de una mayor cantidad de información reguladora a un gen o genes diana y por tanto un mayor grado de pleiotropía, produciendo un incremento en el potencial del organismo o especie para generar diversidad y/o aumentar su complejidad morfológica y funcional (Fig. D_2). Uno de los momentos más interesantes de la historia evolutiva de los metazoos fue lo que se conoce como explosión Cámbrica (hace aprox. 540 millones de años), en este corto periodo de la historia de la Tierra se diversificaron la mayor parte de los organismos bilaterales (subconjunto más numeroso de los Metazoos). Muchos factores abióticos y ecológicos pudieron tener su papel en esta rápida diversificación morfológica (Canfield y cols 2007, Conway-Morris y cols 2006, Marshall y cols 2006, Budd y cols 2008). Sin embargo la base molecular de la diversificación de los planes corporales de los bilaterales debe radicar en cambios genéticos. Muchos de los genes de desarrollo y moléculas de señalización están también presentes en metazoos no bilaterales, sugiriendo que no hubo grandes innovaciones en estos términos en los organismos bilaterales (Erwin 2009, Marshall y Valentine 2010). Heger y colaboradores (Heger y cols 2011) han sugerido que la proteína CTCF ha podido tener un papel muy importante en esta diversificación de los organismos bilaterales. Argumentan que tanto la proteína CTCF como sus sitios de unión están restringidos a organismos bilaterales. El estudio del complejo *Hox* en distintos organismos bilaterales revela una correlación entre la pérdida de la proteína CTCF y la descomposición del complejo. Esto sugiere que en ausencia de interacciones estructurales mediadas por CTCF (entre otras proteínas arquitectónicas), las interacciones funcionales (potenciador-promotor) de larga distancia son poco eficientes, perdiendo relevancia funcional y eliminando las constricciones reguladoras que mantienen el complejo *Hox* unido. Una de las especies que no posee la proteína CTCF es *C.elegans*, dato que correlaciona con el hecho de que el genoma de este organismo no presenta una organización basada en TADs (Crane y cols 2015). En definitiva estos datos limitados apuntarían a que la adquisición de potencial generador de diversidad morfológica es en parte una consecuencia de la capacidad de generar estructuras 3D de la cromatina más elaboradas que favorezcan la co-localización de muchos elementos genómicos en un espacio común.

En conclusión, la habilidad de elaborar organizaciones 3D de la cromatina complejas y ricas en interacciones favorecería la generación de paisajes reguladores extensos que

proporcionarían una gran versatilidad transcripcional de los genes asociados a dichos paisajes. Esto favorecería la aparición de nuevas y más complejas conexiones de redes de regulación génica durante el desarrollo embrionario de un organismo, promoviendo la generación de diversidad morfológica y funcional sin necesidad de crear nuevos elementos génicos. Esta habilidad dependería en gran medida del número de proteínas arquitectónicas y de sitios de unión a las mismas.

Conclusiones

Conclusiones

CONCLUSIONS

1. The zebrafish *six3a-six2a* cluster is divided into two topological associating domains (TADs), with the topological border located between *six2a* and *six3a* genes. This chromatin organization is detected during all analyzed developmental stages (epiboly to pharyngula) independently of the transcriptional state of both genes.
2. The *six2a* TAD contain cis-regulatory elements with enhancer activity that reproduce the endogenous expression of *six2a* gene, and the *six3a* TAD contain cis-regulatory elements with enhancer activity that reproduce the endogenous expression of *six3a* gene.
3. The zebrafish *six3b-six2b*, *six6a-six1a-six4a* and *six6b-six1b-six4b* clusters display a 3D organization that resembles the one observed at the *six3a-six2a* cluster. They are divided into two adjacent TADs with the border present between *six1/2* and *six3/6* genes. This organization is also detected in all developmental staged analyzed and is independent of the transcriptional activity of the genes within the clusters.
4. The mouse *Six3-Six2* and *Six6-Six1-Six4* clusters are also 3D organized into two adjacent TADs, with the topological border located between *Six3/6* and *Six1/2* genes. In the *Six3-Six2* this organization is present in two different developmental stages, and in both clusters is also detected in mouse embryonic stem cells.
5. In both zebrafish and mouse, the interaction profiles and expression patterns of Six genes located within the same TAD are very similar while the Six genes of the same cluster but located within different TADs display almost non-overlapping expression a chromatin contacts profiles.
6. Sea Urchin *Six* cluster 4C-seq data analysis suggest that the spatial organization in this cluster is structured into two adjacent TADs with a topological border located between the *Sp-six3/6* and *Sp-six1/2* genes. The separation of these two genes by this chromatin organization is also supported by the different expression patterns of *Sp-six3/6* and *Sp-six1/2* during larval development.
7. Our data strongly suggest that the *Six* clusters share a deeply evolutionarily conserved 3D chromatin organization that could predate the Cambrian explosion.
8. Deletion of the region containing the topological border present in the zebrafish *six2a-six3a* cluster using a BAC recombineering approach, produces a partial rewiring of the *six2a* and *six3a* expression domains during zebrafish development. These expression domains changes are accompanied by concordant modifications in contact profiles of cis-regulatory elements associated with each gene.
9. Analysis of zebrafish, mouse and sea urchin *Six* cluster topological domains borders suggest the conserved presence of tandemly arrayed and diverging oriented CTCF binding sites at this locations.
10. Genome-wide bioinformatics analyses in human and mouse suggest that this diverging orientation of CTCF binding sites is a general signature of TAD borders.

Anexos

1. ANEXO 1: Protocolo 4C-seq

A continuación se describe el protocolo de preparación de muestras de 4C desde el momento en que se tratan la células con paraformaldeído para la fijación de las interacciones proteína-proteína y proteína-ADN, así como el proceso de PCRs y preparación de librerías.

Fixation and cell lysis

All the steps in this protocol are optimized for using 1×10^7 cells.

1. Add 5ml PFA 4% (2% final concentration) and incubate 10min @ RT while tumbling.
2. Add 1.425ml 2M glycine, mix and put tubes immediately on ice to quench the cross-linking reaction.
3. Centrifuge 8' 1300 rpm at 4°C. Remove supernatant and resuspend gently in 5ml Cold Lysis Buffer (I usually start with the 3C recipe and add NP-40 and Triton if necessary). Mix 3µl of cells with 3µl of Methyl Green-Pyronin staining (Sigma #HT70116) on a microscope slide and overlay with a coverslip. Assess the lysis efficiency using a microscope. Cytoplasm stains pink and the nuclei stains blue/green. When cell lysis is incomplete, douncing can be applied to increase efficiency.
Note: cell lysis is an important step in the protocol as failure of lysis can hamper digestion efficiency.
4. Centrifuge 5min 1800rpm @4°C and remove supernatant.
At this point nuclei can be stored for later use (freeze the pellet in liquid nitrogen and store at -80°C) or resuspended in 450ul Milli-Q and proceed with step 5).

1st Digestion: DpnII

5. Add 60µl of 10X DpnII restriction Buffer.
6. Place the tube @37°C and add 15 µl 10% SDS.
7. Incubate 1hr @37°C while shaking at 900 RPM using an Eppendorf Thermomixer.
8. Add 75µl 20% Triton X-100.
9. Incubate 1hr @37°C while shaking at 900 RPM.
10. Take a 5µl aliquot of the sample as the “Undigested control” and store at 4°C until used in step 14.
11. Add 400U *DpnII*; incubate O/N at 37°C while shaking at 900 RPM.
12. Add 200U *DpnII*; incubate 4 hrs at 37°C while shaking at 900 RPM.
13. Take a 5µl aliquot of the sample as the “Digested control”.
14. Determination of the digestion efficiency (Meanwhile leave digestion running):
 - 14.1. Add 90µl 10mM Tris-HCl pH 7.5 to the 5µl samples from step 10 and 13.
 - 14.2. Add 5µl Prot K (10 mg/ml) and incubate for 1 hour @65°C.
 - 14.3. Add 100µl Phenol-Chloroform (Sigma) to the samples and mix vigorously.
 - 14.4. Spin @13000rpm for 5 min at RT.
 - 14.5. Transfer water phase to a clean tube and load ~ 20µl on a 0.6% agarose gel.
Alternatively, Q-PCR analysis can be used for more precise determination digestion efficiency using multiple primer sets spanning a restriction site. This step is highly recommended when 4C is applied for the first time.
 - 14.6. If digestion is OK proceed with step 15, otherwise repeat from step 11.

1st Ligation

15. Heat-inactivate the restriction enzyme by incubating 20 min. at 65°C and continue with **step 16**. Alternatively, when the restriction enzyme is not sensitive to heat inactivation, e.g. *Bgl*II, continue with **step 15.1**.
 - 15.1. Add 80µl 10% SDS and incubate 30 min. at 65°C.
 - 15.2. Transfer the sample to a 50ml Falcon tube and add 5.4ml Milli-Q
 - 15.3. Add 700µl 10X Ligase buffer Add 375µl 20% TX-100 and incubate 1hr 37°C
 - 15.4. Continue with **step 19**.
16. Transfer the sample to a 50ml Falcon tube.
17. Add 5.7ml Milli-Q.
18. Add 700µl 10X Ligase buffer.
19. Add 60U T4 DNA Ligase mix by swirling and incubate O/N at 16°C. (Shaking is not necessary)
20. Take a 100µl aliquot of the sample as the “Ligation control”.
21. Determine ligation efficiency: (Meanwhile leave ligation running).
 - 21.1. Add 5µl Prot K (10mg/ml) and incubate for 1 hour @65°C.
 - 21.2. Add 100µl Phenol-Chloroform to the sample and mix vigorously.
 - 21.3. Spin 10 min, 16400g at RT.
 - 21.4. Transfer water phase to a clean tube and load ~ 20µl on a 0.6% agarose gel next to the ‘digestion control’ **from step 14**.
 - 21.5. If ligation is OK, proceed with **step 22**. If not, add fresh ATP (final concentration of 1mM) and repeat from **step 19**.

Reverse cross-linking and precipitation

22. Add 30µl Prot K (10mg/ml) and reverse cross-link O/N at 65°C.
23. Add 30µl RNase A (10mg/ml) and incubate 45 minutes at 37°C.
24. Add 7ml Phenol-Chloroform, mix vigorously.
25. Centrifuge 15 min, 3270g at RT.
26. Transfer the aqueous phase to a new 50ml Falcon tube and add: 7.5ml Milli-Q, 1ml 3M NaAC pH 5.6, 7µl Glycogen (1mg/ml, Roche #10901393001) and 35ml 100% EtOH.
Increasing the volume twice before precipitation (partially) prevents the co-precipitation of DTT from the ligase buffer and therefore results in a sample with higher purity.
27. Mix and incubate at -80°C until the sample is frozen solid. (This takes several hours)
28. Spin 60 min, 4000rpm @4°C.
29. Remove supernatant and add 10 ml cold 70% ethanol.
30. Centrifuge 30 min, 4000rpm at 4°C.
31. Remove the supernatant and briefly dry the pellet at RT.
32. Dissolve the pellet in 150µl 10mM Tris-HCl pH 7.5 @37°C.
33. Continue with **step 34** or store sample at -20°C.

Second Digestion: Csp6I

34. To 150µl 3C sample (~1x10⁷ cells) add:
 - 50µl 10X *Csp6I* restriction buffer
 - Milli-Q to 500µl
 - 50U *Csp6I*
35. Incubate O/N @37°C.
36. Take a 5µl aliquot of the sample as the “Digestion control”.
37. Determine digestion efficiency:
 - 37.1. Add 95µl 10mM Tris-HCl pH 7.5 to the 5µl sample from **step 36**.
 - 37.2. Load ~20µl on a 0.6% agarose gel next to the ‘ligation control’ from **step 21**.
 - 37.3. If digestion is OK, proceed with **step 38**. If not, add fresh restriction enzyme and repeat from **step 35**. Alternatively the sample can be re-purified to facilitate efficient digestion.

Second Ligation and purification

- 38.** Inactivate enzyme by incubating @65°C for 25 minutes and continue with step 39.
If not heat sensitive, the restriction enzyme can be inactivated by sample purification (step 38.1).
- 38.1.** Add 500µl Phenol-Chloroform and mix vigorously
- 38.2.** Spin 10 min, 16400g at RT.
- 38.3.** Transfer the aqueous phase to a fresh tube and add 50µl 2M NaAc pH 5.6 and 950µl 100% EtOH
- 38.4.** Incubate at -80°C until completely frozen
- 38.5.** Spin 20min 16400g at 4°C
- 38.6.** Remove supernatant and add 150µl cold 70% ethanol.
- 38.7.** Spin 10min 16400g at 4°C
- 38.8.** Resuspend the pellet in 500µl 10mM Tris-HCl pH 7.5
- 39.** Transfer sample to a 50ml tube and add:
 - 12.1ml Milli-Q
 - 1.4 ml 10X Ligation buffer
 - 100U T4 DNA Ligase
- 40.** Ligate O/N at 16°C.
- 41.** Add: 466µl 3M NaAC pH 5.6, 7µl Glycogen (1mg/ml) and 35ml 100% EtOH. Mix well.
- 42.** Store at -80°C until completely frozen.
- 43.** Spin 1h 30 min, 4000rpm @4°C
- 44.** Remove the supernatant and add 10ml cold 70% ethanol.
- 45.** Spin 30 min, 4000rpm @4°C.
- 46.** Remove supernatant and dry the pellet @RT.
- 47.** Dissolve the pellet in 150µl 10mM Tris-HCl pH 7.5 @37°C.
- 48.** Purify samples with the QIAquick PCR purification kit (Qiagen #28104)
Use 3 columns per sample; binding capacity is 10µg DNA per column.
Elute columns with 50µl 10mM Tris-HCl pH 7.5 and pool samples.
- 49.** Measure concentration using the Nanodrop spectrophotometer and run a serial dilution of 0.125, 0.25, 0.5 and 1µl sample on a 2% agarose gel in order to estimate the concentration compared to a reference sample, e.g. phage-λ DNA.
- 50.** The 4C template is now finished and can be stored at -20°C or continued directly in step 51.

PCR

- 51.** Determine linear range of amplification by performing a PCR using template dilutions of 12.5, 25, 50 and 100ng 4C template. A typical 25µl PCR reaction consist of:
 - 2.5µl 10X PCR buffer 1(supplied with the Expand Long Template Polymerase)
 - 0.5µl dNTP (10mM)
 - 35pmol forward primer (1.5µl of a 1/7 dilution from a 1µg/µl 20nt primer stock)
 - 35pmol reverse primer (1.5µl of a 1/7 dilution from a 1µg/µl 20nt primer stock)
 - 0.35µl Expand Long Template Polymerase (Roche #11759060001)
 - X µl Milli-Q to a total volume of 25µl

A typical 4C-PCR program: 2' 94 °C; 10" 94 °C; 1' 55 °C; 3' 68 °C; 30x repeat; 5' 68 °C; ∞ 12 °C. The concentration of primers used in a 4C-PCR is typically three times higher than a regular PCR as this often facilitates the efficiency of amplification.
- 52.** Separate 15µl PCR product on a 1.5% agarose gel and quantify to assess linear amplification and template quality.
- 53.** Determine the functionality of the adaptor primers by comparing them with the 'short' primers from step 51. Note the volume of the adaptor primers is corrected for their length difference by using 4.5µl and 3µl of a 1/7 diluted 1µg/µl stock solution of the ~75nt reading primer and the ~40nt reverse primer respectively. The adaptor primers should cause a shift in PCR product length, which should be visible when separated and compared on a 1.5% agarose gel.
- 54.** When satisfied about the quality and quantity of the PCR product generated using the adaptor primers, the high complexity PCR is performed.

- 80µl 10X PCR buffer 1
- 16µl dNTP (10mM)
- 1.12nmol 75nt reading primer (24µl reading primer of a 1µg/µl 75nt primer stock)
- 1.12nmol 40nt reverse primer (16µl reverse primer of a 1µg/µl 40nt primer stock)
- typically 3.2µg 4C template
- 11.2µl Expand Long Template polymerase
- Milli-Q water till 800ul total

Mix and separate into 16 reactions of 50µl before running the PCR

- 55.** Collect and pool the 16 reactions. Purify the sample using the High Pure PCR Product Purification Kit (Roche #11732676001), which effectively separates between the non-used adaptor primers (~75nt) and the PCR product (>120nt). Use minimal two columns per 16 reactions. We normally use three columns per sample.
- 56.** Determine sample quantity and purity using the Nanodrop-spectrophotometer. Typically the yield resides between 10 and 20µg with A260/A280 ~1.85 and A260/A230 >1.5. Control of sample purity is important in order to prevent complications during the sequencing procedure. If absorption ratios deviate re-purification is advised. We also check concentration using picogreen measurement.
- 57.** Quality is determined by separation of 300ng purified PCR product on a 1.5% agarose gel.
- 58.** Combine 4C PCR products of different experiments in preferred ratios for sequencing. We send the samples separately.

10X Ligase buffer_Promega:

Final concentration	In the lab we have	Volume
300mM Tris-HCl pH7.8	1M pH 8	3ml
100mM MgCl ₂	1M	1ml
100mM DTT	Powder	0.1543g
10mM ATP	100mM	100ul
H ₂ O		ml
	Final volume	10ml

2. ANEXO 2: Protocolo ChIP-4C.

A continuación se describen los pasos específicos que hay que añadir o cambiar al protocolo de 4C-seq del anexo uno para llevar a cabo un ChIP-4C. Este protocolo se trata de una “prueba de principio” que no ha sido validada y que probablemente requiera muchos cambios y optimizaciones.

- 1) Respecto a la primera digestión, se añaden un total de 200 unidades más de DpnII. A pesar de incrementar en 10 veces la cantidad de muestra, el aumento de solo un 25% de la cantidad de enzima fue suficiente para digerir eficazmente la muestra. Esto probablemente indica que usamos enzima en exceso.
16. Add **250U** DpnII; incubate 4 hrs at 37°C while shaking at 900 RPM.
17. Add **300U** DpnII; incubate O/N at 37°C while shaking at 900 RPM.
18. Add **250U** DpnII; incubate 4 hrs at 37°C while shaking at 900 RPM.
- 2) Respecto a la segunda ligación añadimos también un 25% más de ligasa.
25. Add **80U** T4 DNA Ligase mix by swirling and incubate O/N at 16°C
- 3) Tras comprobar que la primera ligación ha funcionado, procedemos al paso de inmunoprecipitación. Estamos pues en el paso 14 del protocolo de 4C, con un falcón de 50 ml a 16°C, conteniendo la reacción de 7ml de ligación. A continuación incluimos estos nuevos pasos
59. To 7 ml o ligation ON mixture add 3ml of IP-4C-Chip-Buffer(10mM Tris pH 8, 77,76mM KCl, 5,56mM EDTA, 7,2mM DTT, 222mM NaCl, 11,1% Glycerol, 2,3% NP40). divide in 6 tube of 2ml.(1700ul per tube)
60. Centrifuge 14000 rpm 10` 4°C. Pass the supernatant to new tubes.
61. Add 1,5ul of Ab H3K27ac or 2ul of Ab H3K27me3 to each tube. Incubate O/N rotating 4°C.

Next Day

62. Wash 120ul of Magnetic Beads in IP Buffer.(20ul of Beads per 1 ml of IP buffer).
63. Resuspend beads in 300ul of IP Buffer.
64. Add 50ul of Beads to each chromatine tube and incubate at 4°C 1 hour.
65. Wash beads once with 1 ml of each of the following buffers (sequentially, everything on ice):
 - ChIP1 buffer
 - ChIP2 buffer
 - ChIP3 buffer
 - ChIP1 buffer.
66. Wash beads once with 0.5 ml TE (everything on ice)
67. Elute in 2x 100 ul NaHCO₃ pH 8.8 (0.1 M) / SDS (1%) 15 min at room temperature. An pool everything in a 2ml tube.
68. Decrosslinking. Add 60ul NaCl 4M and 10 protK 10mg/ml. O/N 65°C.

Next Day

69. Add 23ul 10mg/ml RNAase 2h 37°C
70. Split sample in 3 eppendorff (1200/3 = 400)
71. Phenolize in 1volume of phenol:chlo:IAA= 400ul
72. Precipitate 20ul 1ug/ul Glycogen + 1/10 Volume AcNa (40ul) 3M + 2V 100% ethanol(1000ul). - 80°C (Until completely frozen).

- 73.** 30´ 14000 rpm. 500ul EtOH 70% and spin 15´ 14000 rpm. Resuspend in 50ul 10mM Tris HCl pH 7.5. Pool 3 tubes in only one.

IP buffer

50 mM Tris pH 8
100 mM NaCl
2 mM EDTA
1 mM DTT
1% NP40
Protease inhibitors

ChIP 1 buffer

IP buffer plus 0.1% deoxycholate

ChIP 2 buffer

IP buffer plus:
400 mM NaCl (extra, 500 mM final)
0.1% deoxycholate

ChIP 3 buffer

IP buffer plus:
250 mM LiCl
0.1% deoxycholate

- 4)** A continuación pasamos al paso 34 del protocolo de 4C y lo continuamos hasta el final sin modificación.

3. ANEXO 3: Cebadores 4C-seq

	Nombre del Cebador	Secuencia	Posición Cebador DpnII
Pez Cebra	six2a_DpnII	GAAGAGAGGCACAAAACCTTTAGATC	chr13:9797970
	six2a_Csp6I	GTCATCGCTTAGATAGACATACAAGTAC	
	six3a_DpnII	AGTGGGTGGAATATTATTGTGATC	chr13:9826837
	six3a_Csp6I	AAGTGGTGAAAGCCTCTACGTAC	
	six2b_DpnII	GCTTACTACTAAGTAAAGTTTTTGA	chr12:27140158
	six2b_Csp6I	TCTCGTCACAACTATCAATCC	
	six3b_DpnII	CCAGAAGCAGAGGGCGA	chr12:27128597
	six3b_Csp6I	CCCACTATACATTTAACTATGC	
	six4a_DpnII	GCGCTGCAGAGTGCATTGATC	chr13:31871941
	six4a_Csp6I	TGTTGTCAAGATTGGGAATAGGGACC	
	six1a_DpnII	ACTTCCGTGAGCTCTACAAGATC	chr13:31845355
	six1a_Csp6I	GGCTCCCCATGCAATCCAC	
	six6a_DpnII	CAATGTTGCCAAACACACGAAGATC	chr13:31826506
	six6a_Csp6I	GCCCTATACGCCAACTTCAAGTC	
	six4b_DpnII	CCCACGACTCTCCCTCTTGATC	chr20:20557180
	six4b_Csp6I	TCGTCTCTGCAAGGATATGTGTAC	
	six1b_DpnII	GTCTCTCCCGCTTGCGATC	chr20:20572332
	six1b_Csp6I	ACTGTGACTCATGTGCGCG	
	six6b_DpnII	TGAGCTGTCAGATGTCTACGAGATC	chr20:20606589
	six6b_Csp6I	GCCCTATACGCCAACTTCAAGTC	
	ppm1ba_DpnII	GCTGATATGCATGCAAGA	chr13:10292253
	ppm1ba_Csp6I	agaattaagcaatatgctagt	
	skc3a1_DpnII	CCGAATTCAACCGAAAGA	chr13:10186394
	skc3a1_Csp6I	CACTGTGCTATTATGGCTATCG	
	prepl_DpnII	GAGAGTGAATACACGCAGA	chr13:10237450
	prepl_Csp6I	TAGGCATCCAGCATACGGCT	
	tm9sf3_DpnII	CGTAAAATTGTTACGCGAGTGATC	chr13:9240919
	tm9sf3_Csp6I	CCTCTCGATGTGTTGGTGATC	
	prdx3_DpnII	GAAGGTGAAGTCTGTTCAAAGATC	chr13:9208662
	prdx3_Csp6I	CATAAAGGTGCTGCTGACGTAC	
	EnhII(R5)_DpnII	TACTCTCAGAGCTGTTAAAGGATC	chr13:9877427
	EnhII(R5)_Csp6I	GTTGCTGCATCTTCTGGAC	
	EnhVI(2.2)_DpnII	GGAGCCACGCGGATTAAGAGATC	chr13:9763732
	EnhVI(2.2)_Csp6I	GATAGCCACAGGAATTCGGCTG	
	six2a:GFP_DpnII	AATGTGGTATGGCTGATTATGATC	chr13:9798124
	six2a:GFP_Csp6I	GGGGCTGTGGGAAGTAC	
	six3a:mCherry_DpnII	cagcttcagctctgctgac	chr13:9826542
	six3a:mCherry_Csp6I	actacaccatctggaacagtac	
	Nombre del Cebador	Secuencia	Posición Cebador DpnII
Ratón	six2_DpnII	CGACTCTGAGTCACAACGATC	chr17:86087424
	six2_Csp6I	CACTACATCGAGGCGGAGAAGC	
	six3_DpnII	GCGCCCTCTGCGTAGAGATC	chr17:86022535
	six3_Csp6I	CCAGCAACTGTCAGCAGCCG	
	six4_DpnII	GTCCCTGCCCCAGAGCgate	chr12:74213772
	six4_Csp6I	TGCCTGCCCAGAAGTTCCGAG	
	six1_DpnII	GGAATCCCTTCTCACTTGGATC	chr12:74149274
	six1_Csp6I	GGGGACTTATACGGGCTCTC	
	six6_DpnII	ACAGGGAGGGGAAGTGGATC	chr12:74040426
	six6_Csp6I	CCAGGAGGCAGAGAAGCTGC	
Erizo de Mar	Sp_six4/5_DpnII	CGCCAATTGAAACGGTGGGATC	Scaffold143:692728
	Sp_six4/5_Csp6I	AAAGTTCGACGGCTTTACATC	
	Sp_six1/2_DpnII	CAAAAATAGGCGACAGCGAGATC	Scaffold143:780065
	Sp_six1/2_Csp6I	AGGATAGGGTTGTGGAGTAC	
	Sp_six3/6_DpnII	GCAATGGCAACTCCTCTTTGAtc	Scaffold143:1022703
	Sp_six3/6_Csp6I	TTGGCCCCGTCGACAAAGTAC	

4. ANEXO 4: Regiones usadas para testar actividad potenciadora

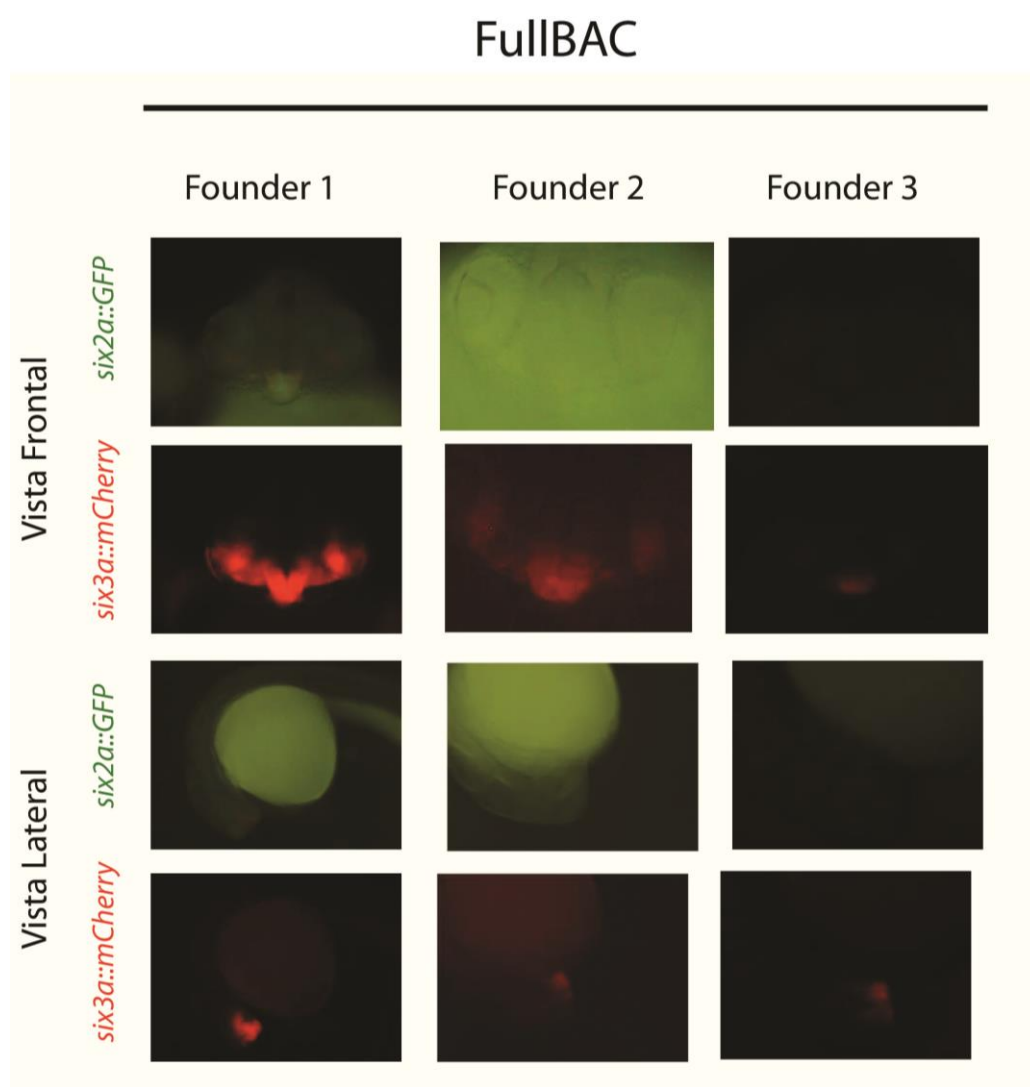
ELEMENTO	CEBADOR FWD	CEBADOR REV	REGIÓN AMPLIFICADA
E-I	AGGAAAGTGCATTACAATTATG	TTTAATAAAAATGTTTCAGTTATTG C	10030695- 10033121
E-II	TGAACACGAGAAGAACATGC	AAGTGACTTTTGTTCCTTCAGG	9895167- 9897443
E-III	TTTGATTCGCCTCATCTTG	TGTGAACATTTTCCTGCTCTG	9876137- 9877960
E-IV	TGCCTCGACTAGGCTACACC	TCCTGTACAGCAGTCCCACA	9844141- 9846528
E-V	AAATTATTTTCGTTTCCAGTCG	AAGTTTTCGACGAACAAATGC	9794569- 9796315
E-VI	GCATGTGGATTTGGGAAACT	TTGTGAGTCTGGTGCTACCG	9763051- 9764894
N1	CAAGCAAAGCTGTTCTCTCT	AGCTCAAGTGCCCTCATCAT	9940230- 9941034
N2	CATGATGTTTTCATCCAAGG	AGAAGATGGACAAATGAGAACC	9903851- 9905652
N3	GAACAATTGCATCATAGCAAAA AC	CAAACAGGTCCCCAAGACAC	9838731- 9840512

5. ANEXO 5: Cebadores para la preparación de sondas de los genes Six de pez cebra

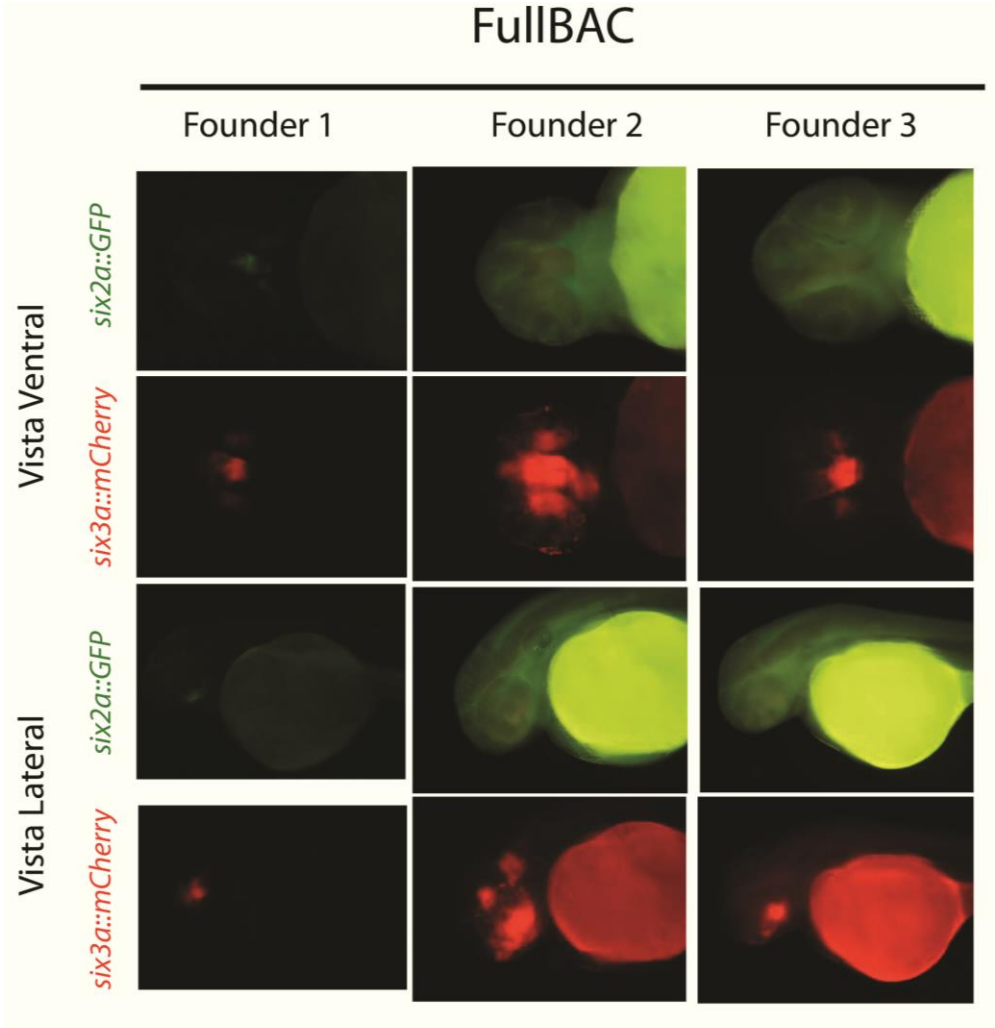
Gen	CEBADOR FWD	CEBADOR REV
<i>six4a</i>	ATGTCTGTTTCCTCTAGTGAAGTG	CCGTACTGTGATTTCCATCCG
<i>six1a</i>	ATGTCAATCTTGCCCTCGTTCG	CTACGATCCTAAATCCACAAGGCTG
<i>six6a</i>	ATGTTCCAGCTGCCCATTTTG	TTAGATGTCACATTCAGTGTGCG
<i>six4b</i>	ATGTGCTGCCTGAAACCCAG	TCTGTACTGTGATTGCCATCCG
<i>six1b</i>	ATGTCAATGTTGCCTTCTTTCG	TTAAGAGCCTAGGTCCACTAGAC
<i>six6b</i>	ATGTTTCAGTTGCCAATTTTGAA	TTAGATGTCACATTCGCTGTC

6. ANEXO 6: Líneas transgénicas estables de las versiones del BAC DKEY-74B2, FullBAC (24hpf, 48hpf, 5dpf) y $\Delta 18\text{kbBAC}$ (5 dpf).

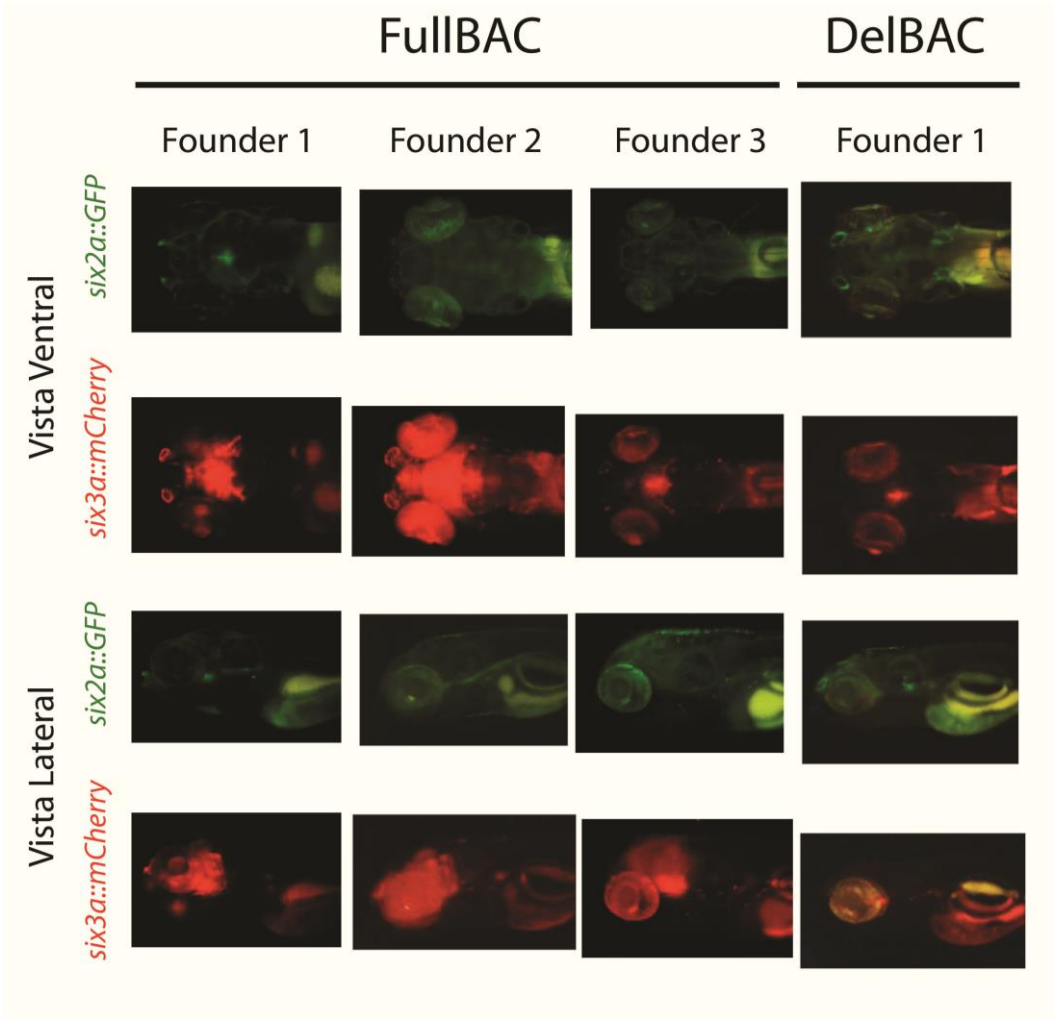
A) 24 hpf



B) 48 hpf



C) 5 dpf



Bibliografía

Bibliografía

Bibliografía

An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57-74. PubMed PMID: 22955616; NIHMSID: NIHMS381381; PubMed Central PMCID: PMC3439153.

Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, François KJ, et al. A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Dev Cell*. 2009 Sep;17(3):425-34. PubMed PMID: 19758566; NIHMSID: NIHMS141694; PubMed Central PMCID: PMC2746918.

Alonso ME, Pernaute B, Crespo M, Gómez-Skarmeta JL, Manzanares M. Understanding the regulatory genome. *Int J Dev Biol*. 2009;53(8-10):1367-78. PubMed PMID: 19247937.

Amano T, Sagai T, Tanabe H, Mizushima Y, Nakazawa H, et al. Chromosomal dynamics at the *Shh* locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell*. 2009 Jan;16(1):47-57. PubMed PMID: 19097946.

Ando H, Kobayashi M, Tsubokawa T, Uyemura K, Furuta T, et al. *Lhx2* mediates the activity of *Six3* in zebrafish forebrain growth. *Dev Biol*. 2005 Nov 15;287(2):456-68. PubMed PMID: 16226737.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, et al. A switch between topological domains underlies *HoxD* genes collinearity in mouse limbs. *Science*. 2013 Jun 7;340(6137):1234167. PubMed PMID: 23744951.

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*. 2002 Aug 23;297(5585):1301-10. PubMed PMID: 12142439.

Banerji J, Olson L, Schaffner W. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*. 1983 Jul;33(3):729-40. PubMed PMID: 6409418.

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science*. 2012 Dec 21;338(6114):1587-93. PubMed PMID: 23258890.

Beaster-Jones L. Cis-regulation and conserved non-coding elements in amphioxus. *Brief Funct Genomics*. 2012 Mar;11(2):118-30. PubMed PMID: 22402505.

Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*. 1999 Aug 6;98(3):387-96. PubMed PMID: 10458613.

Bessa J, Tena JJ, de la Calle-Mustienes E, Fernández-Miñán A, Naranjo S, et al. Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev Dyn*. 2009 Sep;238(9):2409-17. PubMed PMID: 19653328.

Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Res*. 2012 Oct;22(10):2043-53. PubMed PMID: 22593555; PubMed Central PMCID: PMC3460198.

Bogdanović O, Fernández-Miñán A, Tena JJ, de la Calle-Mustienes E, Gómez-Skarmeta JL. The developmental epigenomics toolbox: ChIP-seq and MethylCap-seq profiling of early zebrafish embryos. *Methods*. 2013 Aug 15;62(3):207-15. PubMed PMID: 23624103.

Bouwman BA, de Laat W. Architectural hallmarks of the pluripotent genome. *FEBS Lett*. 2015 Oct 7;589(20 Pt A):2905-13. PubMed PMID: 25957773.

Boyle S, Rodesch MJ, Halvensleben HA, Jeddloh JA, Bickmore WA. Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res*. 2011 Oct;19(7):901-9. PubMed PMID: 22006037; PubMed Central PMCID: PMC3210351.

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, et al. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet*. 2001 Feb 1;10(3):211-9. PubMed PMID: 11159939.

Bibliografía

Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol.* 1971 Jun;46(2):111-38. PubMed PMID: 5160087.

Brown JM, Green J, das Neves RP, Wallace HA, Smith AJ, et al. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol.* 2008 Sep 22;182(6):1083-97. PubMed PMID: 18809724; PubMed Central PMCID: PMC2542471.

Budd GE. The earliest fossil record of the animals and its significance. *Philos Trans R Soc Lond B Biol Sci.* 2008 Apr 27;363(1496):1425-34. PubMed PMID: 18192192; PubMed Central PMCID: PMC2614223.

Buecker C, Wysocka J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* 2012 Jun;28(6):276-84. PubMed PMID: 22487374.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013 Dec;10(12):1213-8. PubMed PMID: 24097267; NIHMSID: NIHMS554473; PubMed Central PMCID: PMC3959825.

Burke LJ, Hollemann T, Pieler T, Renkawitz R. Molecular cloning and expression of the chromatin insulator protein CTCF in *Xenopus laevis*. *Mech Dev.* 2002 Apr;113(1):95-8. PubMed PMID: 11900981.

Bussmann J, Schulte-Merker S. Rapid BAC selection for tol2-mediated transgenesis in zebrafish. *Development.* 2011 Oct;138(19):4327-32. PubMed PMID: 21865323.

Cañestro C, Albalat R, Irimia M, Garcia-Fernández J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin Cell Dev Biol.* 2013 Feb;24(2):83-94. PubMed PMID: 23291262.

Canfield DE, Poulton SW, Narbonne GM. Late-Neoproterozoic deep-ocean oxygenation and the rise of animal life. *Science.* 2007 Jan 5;315(5808):92-5. PubMed PMID: 17158290.

Cavalli G, Misteli T. Functional implications of genome topology. *Nat Struct Mol Biol.* 2013 Mar;20(3):290-9. PubMed PMID: 23463314.

Chen K, Johnston J, Shao W, Meier S, Staber C, et al. A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *Elife.* 2013 Aug 13;2:e00861. PubMed PMID: 23951546; PubMed Central PMCID: PMC3743134.

Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* 2012 Mar;22(3):490-503. PubMed PMID: 22270183; PubMed Central PMCID: PMC3292289.

Choi S, Kim UJ. Construction of a bacterial artificial chromosome library. *Methods Mol Biol.* 2001;175:57-68. PubMed PMID: 11462856.

Ciotto G, Hofemeister H, Maresca M, Fu J, Sarov M, et al. Recombineering BAC transgenes for protein tagging. *Methods.* 2011 Feb;53(2):113-9. PubMed PMID: 20868752.

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002 Feb;9(2):279-89. PubMed PMID: 11864602.

Conway Morris S. Darwin's dilemma: the realities of the Cambrian 'explosion'. *Philos Trans R Soc Lond B Biol Sci.* 2006 Jun 29;361(1470):1069-83. PubMed PMID: 16754615; PubMed Central PMCID: PMC1578734.

Copley RR. The animal in the genome: comparative genomics and evolution. *Philos Trans R Soc Lond B Biol Sci.* 2008 Apr 27;363(1496):1453-61. PubMed PMID: 18192189; PubMed Central PMCID: PMC2614226.

Cramer P. Common structural features of nucleic acid polymerases. *Bioessays.* 2002 Aug;24(8):724-9. PubMed PMID: 12210533.

Bibliografía

- Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*. 2015 Jul 9;523(7559):240-4. PubMed PMID: 26030525; NIHMSID: NIHMS679137; PubMed Central PMCID: PMC4498965.
- Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, et al. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc Natl Acad Sci U S A*. 2004 Jan 27;101(4):992-7. PubMed PMID: 14732688; PubMed Central PMCID: PMC327130.
- Cremer T, Cremer C. Rise, fall and resurrection of chromosome territories: a historical perspective Part I The rise of chromosome territories. *Eur J Histochem*. 2006 Jul-Sep;50(3):161-76. PubMed PMID: 16920639.
- Cremer T, Cremer C. Rise, fall and resurrection of chromosome territories: a historical perspective Part II Fall and resurrection of chromosome territories during the 1950s to 1980s Part III Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. *Eur J Histochem*. 2006 Oct-Dec;50(4):223-72. PubMed PMID: 17213034.
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, et al. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*. 1999 Jun 14;145(6):1119-31. PubMed PMID: 10366586; PubMed Central PMCID: PMC2133153.
- Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*. 2009 Jan;19(1):24-32. PubMed PMID: 19056695; PubMed Central PMCID: PMC2612964.
- Cullen KE, Kladde MP, Seyfred MA. Interaction between transcription regulatory regions of prolactin chromatin. *Science*. 1993 Jul 9;261(5118):203-6. PubMed PMID: 8327891.
- Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. *Biochim Biophys Acta*. 2015 Aug;1849(8):1116-31. PubMed PMID: 25934543.
- Davidson EH, Erwin DH. Evolutionary innovation and stability in animal gene networks. *J Exp Zool B Mol Dev Evol*. 2010 May 15;314(3):182-6. PubMed PMID: 19937660.
- de la Calle-Mustienes E, Feijóo CG, Manzanares M, Tena JJ, Rodríguez-Seguel E, et al. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res*. 2005 Aug;15(8):1061-72. PubMed PMID: 16024824; PubMed Central PMCID: PMC1182218.
- de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*. 2013 Oct 24;502(7472):499-506. PubMed PMID: 24153303.
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002 Feb 15;295(5558):1306-11. PubMed PMID: 11847345.
- Dekker J, Heard E. Structural and functional diversity of Topologically Associating Domains. *FEBS Lett*. 2015 Oct 7;589(20 Pt A):2877-84. PubMed PMID: 26348399; NIHMSID: NIHMS721424; PubMed Central PMCID: PMC4598308.
- Dickel DE, Visel A, Pennacchio LA. Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1620):20120359. PubMed PMID: 23650633; PubMed Central PMCID: PMC3682724.
- Ding N, Zhou H, Esteve PO, Chin HG, Kim S, et al. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Mol Cell*. 2008 Aug 8;31(3):347-59. PubMed PMID: 18691967; NIHMSID: NIHMS65230; PubMed Central PMCID: PMC2583939.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012 Apr 11;485(7398):376-80. PubMed PMID: 22495300; NIHMSID: NIHMS366885; PubMed Central PMCID: PMC3356448.
- Dong X, Navratilova P, Fredman D, Drivenes Ø, Becker TS, et al. Exonic remnants of whole-genome duplication reveal cis-regulatory function of coding exons. *Nucleic Acids Res*. 2010 Mar;38(4):1071-85. PubMed PMID: 19969543; PubMed Central PMCID: PMC2831330.

Bibliografía

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 2006 Oct;16(10):1299-309. PubMed PMID: 16954542; PubMed Central PMCID: PMC1581439.

Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLf. *Genes Dev.* 2004 Oct 15;18(20):2485-90. PubMed PMID: 15489291; PubMed Central PMCID: PMC529536.

Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods.* 2012 Feb 28;9(3):215-6. PubMed PMID: 22373907; NIHMSID: NIHMS441097; PubMed Central PMCID: PMC3577932.

Erokhin M, Vassetzky Y, Georgiev P, Chetverina D. Eukaryotic enhancers: common features, regulation, and participation in diseases. *Cell Mol Life Sci.* 2015 Jun;72(12):2361-75. PubMed PMID: 25715743.

Erwin DH. Early origin of the bilaterian developmental toolkit. *Philos Trans R Soc Lond B Biol Sci.* 2009 Aug 12;364(1527):2253-61. PubMed PMID: 19571245; PubMed Central PMCID: PMC2873006.

Faure AJ, Schmidt D, Watt S, Schwalie PC, Wilson MD, et al. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* 2012 Nov;22(11):2163-75. PubMed PMID: 22780989; PubMed Central PMCID: PMC3483546.

Ferraiuolo MA, Rousseau M, Miyamoto C, Shenker S, Wang XQ, et al. The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* 2010 Nov;38(21):7472-84. PubMed PMID: 20660483; PubMed Central PMCID: PMC2995065.

Filippova GN, Lindblom A, Meincke LJ, Klenova EM, Neiman PE, et al. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer.* 1998 May;22(1):26-36. PubMed PMID: 9591631.

Fogelgren B, Kuroyama MC, McBratney-Owen B, Spence AA, Malahn LE, et al. Misexpression of Six2 is associated with heritable frontonasal dysplasia and renal hypoplasia in 3H1 Br mice. *Dev Dyn.* 2008 Jul;237(7):1767-79. PubMed PMID: 18570229; NIHMSID: NIHMS225974; PubMed Central PMCID: PMC2955765.

Fraser J, Rousseau M, Shenker S, Ferraiuolo MA, Hayashizaki Y, et al. Chromatin conformation signatures of cellular differentiation. *Genome Biol.* 2009;10(4):R37. PubMed PMID: 19374771; PubMed Central PMCID: PMC2688928.

Frith MC, Fu Y, Yu L, Chen JF, Hansen U, et al. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.* 2004;32(4):1372-81. PubMed PMID: 14988425; PubMed Central PMCID: PMC390287.

Fullwood MJ, Ruan Y. ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem.* 2009 May 1;107(1):30-9. PubMed PMID: 19247990; NIHMSID: NIHMS136997; PubMed Central PMCID: PMC2748757.

Gallardo ME, Lopez-Rios J, Feraud-Espinosa I, Granadino B, Sanz R, et al. Genomic cloning and characterization of the human homeobox gene SIX6 reveals a cluster of SIX genes in chromosome 14 and associates SIX6 hemizygosity with bilateral anophthalmia and pituitary anomalies. *Genomics.* 1999 Oct 1;61(1):82-91. PubMed PMID: 10512683.

Gavrilov A, Eivazova E, Priozhkova I, Lipinski M, Razin S, et al. Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol Biol.* 2009;567:171-88. PubMed PMID: 19588093.

Gehrke AR, Schneider I, de la Calle-Mustienes E, Tena JJ, Gomez-Marin C, et al. Deep conservation of wrist and digit enhancers in fish. *Proc Natl Acad Sci U S A.* 2015 Jan 20;112(3):803-8. PubMed PMID: 25535365; PubMed Central PMCID: PMC4311833.

Gestri G, Carl M, Appolloni I, Wilson SW, Barsacchi G, et al. Six3 functions in anterior neural plate specification by promoting cell proliferation and inhibiting Bmp4 expression. *Development.* 2005

Bibliografía

May;132(10):2401-13. PubMed PMID: 15843413; NIHMSID: UKMS28203; PubMed Central PMCID: PMC2789257.

Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, et al. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*. 2014 Aug 7;512(7512):96-100. PubMed PMID: 25043061.

Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*. 2014 May 8;157(4):950-63. PubMed PMID: 24813616; NIHMSID: NIHMS686568; PubMed Central PMCID: PMC4427251.

Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*. 2007 Jun;17(6):877-85. PubMed PMID: 17179217; PubMed Central PMCID: PMC1891346.

Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014 Dec;289(6):1045-60. PubMed PMID: 25092473.

Gómez-Díaz E, Corces VG. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol*. 2014 Nov;24(11):703-11. PubMed PMID: 25218583; NIHMSID: NIHMS627751; PubMed Central PMCID: PMC4254322.

Goodrich JA, Tjian R. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet*. 2010 Aug;11(8):549-58. PubMed PMID: 20628347; NIHMSID: NIHMS243534; PubMed Central PMCID: PMC2965628.

Gray CE, Coates CJ. Cloning and characterization of cDNAs encoding putative CTCFs in the mosquitoes, *Aedes aegypti* and *Anopheles gambiae*. *BMC Mol Biol*. 2005 Jun 28;6:16. PubMed PMID: 15985163; PubMed Central PMCID: PMC1174870.

Gregory SG, Barlow KF, McLay KE, Kaul R, Swarbreck D, et al. The DNA sequence and biological annotation of human chromosome 1. *Nature*. 2006 May 18;441(7091):315-21. PubMed PMID: 16710414.

Grob S, Schmid MW, Grossniklaus U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol Cell*. 2014 Sep 4;55(5):678-93. PubMed PMID: 25132176.

Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, et al. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev*. 1996 Jul 1;10(13):1670-82. PubMed PMID: 8682297.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008 Jun 12;453(7197):948-51. PubMed PMID: 18463634.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*. 2007 Jul 13;130(1):77-88. PubMed PMID: 17632057; NIHMSID: NIHMS85637; PubMed Central PMCID: PMC3200295.

Guglielmi B, La Rochelle N, Tjian R. Gene-specific transcriptional mechanisms at the histone gene cluster revealed by single-cell imaging. *Mol Cell*. 2013 Aug 22;51(4):480-92. PubMed PMID: 23973376.

Guo Y, Xu Q, Canzio D, Shou J, Li J, et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. 2015 Aug 13;162(4):900-10. PubMed PMID: 26276636.

Hagège H, Klous P, Braem C, Splinter E, Dekker J, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*. 2007;2(7):1722-33. PubMed PMID: 17641637.

Hardison RC, Taylor J. Genomic approaches towards finding cis-regulatory modules in animals. *Nat Rev Genet*. 2012 Jun 18;13(7):469-83. PubMed PMID: 22705667; NIHMSID: NIHMS429527; PubMed Central PMCID: PMC3541939.

- Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* 2011 Oct;7(10):e1002266. PubMed PMID: 22028662; PubMed Central PMCID: PMC3197655.
- Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell.* 2010 May 7;6(5):479-91. PubMed PMID: 20452322; NIHMSID: NIHMS193263; PubMed Central PMCID: PMC2867844.
- Heck MM. Condensins, cohesins, and chromosome architecture: how to make and break a mitotic chromosome. *Cell.* 1997 Oct 3;91(1):5-8. PubMed PMID: 9335328.
- Heger P, Marin B, Bartkuhn M, Schierenberg E, Wiehe T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc Natl Acad Sci U S A.* 2012 Oct 23;109(43):17507-12. PubMed PMID: 23045651; PubMed Central PMCID: PMC3491479.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet.* 2007 Mar;39(3):311-8. PubMed PMID: 17277777.
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science.* 2002 Oct 4;298(5591):129-49. PubMed PMID: 12364791.
- Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell.* 2012 Nov 9;48(3):471-84. PubMed PMID: 23041285; NIHMSID: NIHMS407878; PubMed Central PMCID: PMC3496039.
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature.* 2013 Apr 25;496(7446):498-503. PubMed PMID: 23594743; NIHMSID: NIHMS488044; PubMed Central PMCID: PMC3703927.
- Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet.* 2014 Feb;46(2):205-12. PubMed PMID: 24413732.
- Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, et al. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* 2012 Dec;22(12):2356-67. PubMed PMID: 22722344; PubMed Central PMCID: PMC3514665.
- Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature.* 2013 Nov 14;503(7475):290-4. PubMed PMID: 24141950; NIHMSID: NIHMS522588; PubMed Central PMCID: PMC3838900.
- Jowett T, Lettice L. Whole-mount in situ hybridizations on zebrafish embryos using a mixture of digoxigenin- and fluorescein-labelled probes. *Trends Genet.* 1994 Mar;10(3):73-4. PubMed PMID: 8178366.
- Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol.* 2010 Mar 15;339(2):225-9. PubMed PMID: 19682982; NIHMSID: NIHMS136955; PubMed Central PMCID: PMC2830304.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010 Sep 23;467(7314):430-5. PubMed PMID: 20720539; NIHMSID: NIHMS225969; PubMed Central PMCID: PMC2953795.
- Kawakami K, Sato S, Ozaki H, Ikeda K. Six family genes--structure and function as transcription factors and their roles in development. *Bioessays.* 2000 Jul;22(7):616-26. PubMed PMID: 10878574.
- Kawakami K, Shima A, Kawakami N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci U S A.* 2000 Oct 10;97(21):11403-8. PubMed PMID: 11027340; PubMed Central PMCID: PMC17212.
- Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet.* 2010 May;11(5):345-55. PubMed PMID: 20376054.

Bibliografía

- King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science*. 1975 Apr 11;188(4184):107-16. PubMed PMID: 1090005.
- Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, et al. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol*. 1993 Dec;13(12):7612-24. PubMed PMID: 8246978; PubMed Central PMCID: PMC364833.
- Kobayashi A, Valerius MT, Mugford JW, Carroll TJ, Self M, et al. Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell Stem Cell*. 2008 Aug 7;3(2):169-81. PubMed PMID: 18682239; NIHMSID: NIHMS53557; PubMed Central PMCID: PMC2561900.
- Kölbl AC, Weigl D, Mulaw M, Thormeyer T, Bohlander SK, et al. The radial nuclear positioning of genes correlates with features of megabase-sized chromatin domains. *Chromosome Res*. 2012 Aug;20(6):735-52. PubMed PMID: 23053570.
- Kumar JP. The sine oculis homeobox (SIX) family of transcription factors as regulators of development and disease. *Cell Mol Life Sci*. 2009 Feb;66(4):565-83. PubMed PMID: 18989625; NIHMSID: NIHMS126087; PubMed Central PMCID: PMC2716997.
- Küpfer K, Kölbl A, Biener D, Dittrich S, von Hase J, et al. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*. 2007 Jun;116(3):285-306. PubMed PMID: 17333233; NIHMSID: UKMS4530; PubMed Central PMCID: PMC2688818.
- Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*. 2006 Jul 11;103(28):10684-9. PubMed PMID: 16815976; PubMed Central PMCID: PMC1484419.
- Lagha M, Bothma JP, Levine M. Mechanisms of transcriptional precision in animal development. *Trends Genet*. 2012 Aug;28(8):409-16. PubMed PMID: 22513408; NIHMSID: NIHMS612469; PubMed Central PMCID: PMC4257495.
- Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature*. 2013 Feb 28;494(7438):497-501. PubMed PMID: 23417068; NIHMSID: NIHMS432226; PubMed Central PMCID: PMC4109059.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921. PubMed PMID: 11237011.
- Le Dily F, Baù D, Pohl A, Vicent GP, Serra F, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev*. 2014 Oct 1;28(19):2151-62. PubMed PMID: 25274727; PubMed Central PMCID: PMC4180976.
- Le Dily F, Beato M. TADs as modular and dynamic units for gene regulation by hormones. *FEBS Lett*. 2015 Oct 7;589(20 Pt A):2885-92. PubMed PMID: 26012375.
- Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A*. 2002 May 28;99(11):7548-53. PubMed PMID: 12032320; PubMed Central PMCID: PMC124279.
- Levine M, Tjian R. Transcription regulation and animal diversity. *Nature*. 2003 Jul 10;424(6945):147-51. PubMed PMID: 12853946.
- Levsky JM, Singer RH. Fluorescence in situ hybridization: past, present and future. *J Cell Sci*. 2003 Jul 15;116(Pt 14):2833-8. PubMed PMID: 12808017.
- Lewis M. Allosterity and the lac Operon. *J Mol Biol*. 2013 Jul 10;425(13):2309-16. PubMed PMID: 23500493.

Bibliografía

- Li Y, Huang W, Niu L, Umbach DM, Covo S, et al. Characterization of constitutive CTCF/cohesin loci: a possible role in establishing topological domains in mammalian genomes. *BMC Genomics*. 2013 Aug 14;14:553. PubMed PMID: 23945083; PubMed Central PMCID: PMC3765723.
- Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, et al. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature*. 2008 Nov 20;456(7220):400-3. PubMed PMID: 18931655; NIHMSID: NIHMS67900; PubMed Central PMCID: PMC2597674.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009 Oct 9;326(5950):289-93. PubMed PMID: 19815776; NIHMSID: NIHMS194459; PubMed Central PMCID: PMC2858594.
- Lim EJ, Lu TX, Blanchard C, Rothenberg ME. Epigenetic regulation of the IL-13-induced human eotaxin-3 gene by CREB-binding protein-mediated histone 3 acetylation. *J Biol Chem*. 2011 Apr 15;286(15):13193-204. PubMed PMID: 21325281; PubMed Central PMCID: PMC3075666.
- Liu W, Lagutin O, Swindell E, Jamrich M, Oliver G. Neuroretina specification in mouse embryos requires Six3-mediated suppression of Wnt8b in the anterior neural plate. *J Clin Invest*. 2010 Oct;120(10):3568-77. PubMed PMID: 20890044; PubMed Central PMCID: PMC2947236.
- Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, et al. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene*. 1990 Dec;5(12):1743-53. PubMed PMID: 2284094.
- Lonfat N, Duboule D. Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Lett*. 2015 Oct 7;589(20 Pt A):2869-76. PubMed PMID: 25913784.
- Losick R. Summary: three decades after sigma. *Cold Spring Harb Symp Quant Biol*. 1998;63:653-66. PubMed PMID: 10384330.
- Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015 May 21;161(5):1012-25. PubMed PMID: 25959774.
- Lutz M, Burke LJ, LeFevre P, Myers FA, Thorne AW, et al. Thyroid hormone-regulated enhancer blocking: cooperation of CTCF and thyroid hormone receptor. *EMBO J*. 2003 Apr 1;22(7):1579-87. PubMed PMID: 12660164; PubMed Central PMCID: PMC152892.
- Lutz M, Burke LJ, Barreto G, Goeman F, Greb H, et al. Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res*. 2000 Apr 15;28(8):1707-13. PubMed PMID: 10734189; PubMed Central PMCID: PMC102824.
- Mahy NL, Perry PE, Gilchrist S, Baldock RA, Bickmore WA. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J Cell Biol*. 2002 May 13;157(4):579-89. PubMed PMID: 11994314; PubMed Central PMCID: PMC2173868.
- Marshall CR. Evolution Fossil record reveals tropics as cradle and museum. *Science*. 2006 Oct 6;314(5796):66-7. PubMed PMID: 17023640.
- Marshall CR, Valentine JW. The importance of preadapted genomes in the origin of the animal bodyplans and the Cambrian explosion. *Evolution*. 2010 May;64(5):1189-201. PubMed PMID: 19930449.
- Martinez-Morales JR. Toward understanding the evolution of vertebrate gene regulatory networks: comparative genomics and epigenomic approaches. *Brief Funct Genomics*. 2015 Aug 20;PubMed PMID: 26293604.
- Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*. 2006;7:29-59. PubMed PMID: 16719718.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012 Sep 7;337(6099):1190-5. PubMed PMID: 22955828; NIHMSID: NIHMS492174; PubMed Central PMCID: PMC3771521.

Bibliografía

- McCollum CW, Amin SR, Pauerstein P, Lane ME. A zebrafish LMO4 ortholog limits the size of the forebrain and eyes through negative regulation of six3b and rx3. *Dev Biol.* 2007 Sep 15;309(2):373-85. PubMed PMID: 17692837; NIHMSID: NIHMS30573; PubMed Central PMCID: PMC2100424.
- McEwen GK, Woolfe A, Goode D, Vavouri T, Callaway H, et al. Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. *Genome Res.* 2006 Apr;16(4):451-65. PubMed PMID: 16533910; PubMed Central PMCID: PMC1457030.
- Meireles-Filho AC, Stark A. Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information. *Curr Opin Genet Dev.* 2009 Dec;19(6):565-70. PubMed PMID: 19913403.
- Merkenschlager M, Odom DT. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell.* 2013 Mar 14;152(6):1285-97. PubMed PMID: 23498937.
- Molodtsova D, Harpur BA, Kent CF, Seevananthan K, Zayed A. Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front Genet.* 2014;5:431. PubMed PMID: 25566318; PubMed Central PMCID: PMC4275039.
- Montavon T, Duboule D. Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. *Trends Cell Biol.* 2012 Jul;22(7):347-54. PubMed PMID: 22560708.
- Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, et al. A regulatory archipelago controls Hox genes transcription in digits. *Cell.* 2011 Nov 23;147(5):1132-45. PubMed PMID: 22118467.
- Moon H, Filippova G, Loukinov D, Pugacheva E, Chen Q, et al. CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep.* 2005 Feb;6(2):165-70. PubMed PMID: 15678159; PubMed Central PMCID: PMC1299244.
- Morey C, Da Silva NR, Perry P, Bickmore WA. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development.* 2007 Mar;134(5):909-19. PubMed PMID: 17251268.
- Müller I, Boyle S, Singer RH, Bickmore WA, Chubb JR. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS One.* 2010 Jul 13;5(7):e11560. PubMed PMID: 20644634; PubMed Central PMCID: PMC2903487.
- Murphy KC. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol.* 1998 Apr;180(8):2063-71. PubMed PMID: 9555887; PubMed Central PMCID: PMC107131.
- Narayanan K, Chen Q. Bacterial artificial chromosome mutagenesis using recombineering. *J Biomed Biotechnol.* 2011;2011:971296. PubMed PMID: 21197472; PubMed Central PMCID: PMC3005948.
- Németh A, Guibert S, Tiwari VK, Ohlsson R, Längst G. Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. *EMBO J.* 2008 Apr 23;27(8):1255-65. PubMed PMID: 18354495; PubMed Central PMCID: PMC2367401.
- Nien CY, Liang HL, Butcher S, Sun Y, Fu S, et al. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genet.* 2011 Oct;7(10):e1002339. PubMed PMID: 22028675; PubMed Central PMCID: PMC3197689.
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. Scanning human gene deserts for long-range enhancers. *Science.* 2003 Oct 17;302(5644):413. PubMed PMID: 14563999.
- Nolis IK, McKay DJ, Mantouvalou E, Lomvardas S, Merika M, et al. Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci U S A.* 2009 Dec 1;106(48):20222-7. PubMed PMID: 19923429; PubMed Central PMCID: PMC2779200.
- Noordermeer D, Branco MR, Splinter E, Klous P, van Ijcken W, et al. Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region. *PLoS Genet.* 2008 Mar 7;4(3):e1000016. PubMed PMID: 18369441; PubMed Central PMCID: PMC2265466.

Bibliografía

- Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, et al. The dynamic architecture of Hox gene clusters. *Science*. 2011 Oct 14;334(6053):222-5. PubMed PMID: 21998387.
- Noordermeer D, de Wit E, Klous P, van de Werken H, Simonis M, et al. Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat Cell Biol*. 2011 Jun 26;13(8):944-51. PubMed PMID: 21706023; NIHMSID: NIHMS314187; PubMed Central PMCID: PMC3151580.
- Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, et al. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife*. 2014 Apr 29;3:e02557. PubMed PMID: 24843030; PubMed Central PMCID: PMC4017647.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012 Apr 11;485(7398):381-5. PubMed PMID: 22495304; NIHMSID: NIHMS433490; PubMed Central PMCID: PMC3555144.
- Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014 Apr;15(4):234-46. PubMed PMID: 24614316; NIHMSID: NIHMS729746; PubMed Central PMCID: PMC4610363.
- Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013 Jun 6;153(6):1281-95. PubMed PMID: 23706625; NIHMSID: NIHMS482418; PubMed Central PMCID: PMC3712340.
- Pope BD, Ryba T, Dileep V, Yue F, Wu W, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 2014 Nov 20;515(7527):402-5. PubMed PMID: 25409831; NIHMSID: NIHMS637449; PubMed Central PMCID: PMC4251741.
- Ptashne M. Gene regulation by proteins acting nearby and at a distance. *Nature*. 1986 Aug 21-27;322(6081):697-701. PubMed PMID: 3018583.
- Pugacheva EM, Kwon YW, Hukriede NA, Pack S, Flanagan PT, et al. Cloning and characterization of zebrafish CTCF: Developmental expression patterns, regulation of the promoter region, and evolutionary aspects of gene organization. *Gene*. 2006 Jun 21;375:26-36. PubMed PMID: 16647825.
- Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 2007 Jul 6;317(5834):86-94. PubMed PMID: 17615350.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011 Feb 10;470(7333):279-83. PubMed PMID: 21160473; NIHMSID: NIHMS350238; PubMed Central PMCID: PMC4445674.
- Ragoczy T, Telling A, Sawado T, Groudine M, Kosak ST. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Res*. 2003;11(5):513-25. PubMed PMID: 12971726.
- Ransick A, Ernst S, Britten RJ, Davidson EH. Whole mount in situ hybridization shows Endo 16 to be a marker for the vegetal plate territory in sea urchin embryos. *Mech Dev*. 1993 Aug;42(3):117-24. PubMed PMID: 8217840.
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014 Dec 18;159(7):1665-80. PubMed PMID: 25497547.
- Roeder RG. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci*. 1996 Sep;21(9):327-35. PubMed PMID: 8870495.
- Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet*. 2013 May;14(5):347-59. PubMed PMID: 23568486; NIHMSID: NIHMS454840; PubMed Central PMCID: PMC4010428.

Bibliografía

Rubio ED, Reiss DJ, Welch PL, Distèche CM, Filippova GN, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*. 2008 Jun 17;105(24):8309-14. PubMed PMID: 18550811; PubMed Central PMCID: PMC2448833.

Sabo PJ, Humbert R, Hawrylycz M, Wallace JC, Dorschner MO, et al. Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc Natl Acad Sci U S A*. 2004 Mar 30;101(13):4537-42. PubMed PMID: 15070753; PubMed Central PMCID: PMC384782.

Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012 Sep 6;489(7414):109-13. PubMed PMID: 22955621; NIHMSID: NIHMS382429; PubMed Central PMCID: PMC3555147.

Schaaf CA, Kwak H, Koenig A, Misulovin Z, Gohara DW, et al. Genome-wide control of RNA polymerase II activity by cohesin. *PLoS Genet*. 2013 Mar;9(3):e1003382. PubMed PMID: 23555293; PubMed Central PMCID: PMC3605059.

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, et al. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*. 2012 Jan 20;148(1-2):335-48. PubMed PMID: 22244452; PubMed Central PMCID: PMC3368268.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012 Feb 3;148(3):458-72. PubMed PMID: 22265598.

Shakes LA, Du H, Wolf HM, Hatcher C, Norford DC, et al. Using BAC transgenesis in zebrafish to identify regulatory sequences of the amyloid precursor protein gene in humans. *BMC Genomics*. 2012 Sep 4;13:451. PubMed PMID: 22947103; PubMed Central PMCID: PMC3546842.

Shen E, Shulha H, Weng Z, Akbarian S. Regulation of histone H3K4 methylation in brain development and disease. *Philos Trans R Soc Lond B Biol Sci*. 2014 Sep 26;369(1652):20140093. PubMed PMID: 25135975; PubMed Central PMCID: PMC4142035.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 2012 Aug 2;488(7409):116-20. PubMed PMID: 22763441; NIHMSID: NIHMS579320; PubMed Central PMCID: PMC4041622.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A*. 1992 Sep 15;89(18):8794-7. PubMed PMID: 1528894; PubMed Central PMCID: PMC50007.

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*. 2006 Nov;38(11):1348-54. PubMed PMID: 17033623.

Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, et al. Obesity-associated variants within FTO form long-range functional connections with IIRX3. *Nature*. 2014 Mar 20;507(7492):371-5. PubMed PMID: 24646999; NIHMSID: NIHMS566189; PubMed Central PMCID: PMC4113484.

Soufi A, Donahue G, Zaret KS. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*. 2012 Nov 21;151(5):994-1004. PubMed PMID: 23159369; NIHMSID: NIHMS419243; PubMed Central PMCID: PMC3508134.

Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*. 2006 Sep 1;20(17):2349-54. PubMed PMID: 16951251; PubMed Central PMCID: PMC1560409.

Stearns FW. One hundred years of pleiotropy: a retrospective. *Genetics*. 2010 Nov;186(3):767-73. PubMed PMID: 21062962; PubMed Central PMCID: PMC2975297.

Summerbell D, Ashby PR, Coutelle O, Cox D, Yee S, et al. The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors. *Development*. 2000 Sep;127(17):3745-57. PubMed PMID: 10934019.

Bibliografía

Suster ML, Abe G, Schouw A, Kawakami K. Transposon-mediated BAC transgenesis in zebrafish. *Nat Protoc.* 2011 Dec 1;6(12):1998-2021. PubMed PMID: 22134125.

Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 2014 Mar;24(3):390-400. PubMed PMID: 24398455; PubMed Central PMCID: PMC3941104.

Symmons O, Spitz F. From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1620):20120358. PubMed PMID: 23650632; PubMed Central PMCID: PMC3682723.

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, et al. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*) Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol.* 1988 Sep 20;203(2):439-55. PubMed PMID: 3199442.

Tena JJ, Alonso ME, de la Calle-Mustienes E, Splinter E, de Laat W, et al. An evolutionarily conserved three-dimensional structure in the vertebrate *Irx* clusters facilitates enhancer sharing and coregulation. *Nat Commun.* 2011;2:310. PubMed PMID: 21556064.

Tena JJ, Neto A, de la Calle-Mustienes E, Bras-Pereira C, Casares F, et al. Odd-skipped genes encode repressors that control kidney development. *Dev Biol.* 2007 Jan 15;301(2):518-31. PubMed PMID: 17011543.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, et al. The accessible chromatin landscape of the human genome. *Nature.* 2012 Sep 6;489(7414):75-82. PubMed PMID: 22955617; NIHMSID: NIHMS377833; PubMed Central PMCID: PMC3721348.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell.* 2002 Dec;10(6):1453-65. PubMed PMID: 12504019.

Torres-Núñez E, Cal L, Suárez-Bregua P, Gómez-Marin C, Moran P, et al. Matricellular protein SPARC/osteonectin expression is regulated by DNA methylation in its core promoter region. *Dev Dyn.* 2015 May;244(5):693-702. PubMed PMID: 25728805.

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell.* 2005 Feb 4;17(3):453-62. PubMed PMID: 15694345.

Venkatesh B, Yap WH. Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. *Bioessays.* 2005 Jan;27(1):100-7. PubMed PMID: 15612032.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. The sequence of the human genome. *Science.* 2001 Feb 16;291(5507):1304-51. PubMed PMID: 11181995.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 2015 Mar 3;10(8):1297-309. PubMed PMID: 25732821; PubMed Central PMCID: PMC4542312.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature.* 2009 Feb 12;457(7231):854-8. PubMed PMID: 19212405; NIHMSID: NIHMS125128; PubMed Central PMCID: PMC2745234.

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. *Nature.* 2009 Sep 10;461(7261):199-205. PubMed PMID: 19741700; NIHMSID: NIHMS225119; PubMed Central PMCID: PMC2923221.

Volpi EV, Bridger JM. FISH glossary: an overview of the fluorescence in situ hybridization technique. *Biotechniques.* 2008 Oct;45(4):385-6, 388, 390 passim. PubMed PMID: 18855767.

Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, et al. A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci U S A.* 2009 Oct 27;106(43):18357-61. PubMed PMID: 19826084; PubMed Central PMCID: PMC2761237.

Bibliografía

- Wamstad JA, Wang X, Demuren OO, Boyer LA. Distal enhancers: new insights into heart development and disease. *Trends Cell Biol.* 2014 May;24(5):294-302. PubMed PMID: 24321408.
- Wang H, Maurano MT, Qu H, Varley KE, Gertz J, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* 2012 Sep;22(9):1680-8. PubMed PMID: 22955980; PubMed Central PMCID: PMC3431485.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature.* 2011 Apr 7;472(7341):120-4. PubMed PMID: 21423168; NIHMSID: NIHMS338997; PubMed Central PMCID: PMC3670758.
- Wapinski OL, Vierbuchen T, Qu K, Lee QY, Chanda S, et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell.* 2013 Oct 24;155(3):621-35. PubMed PMID: 24243019; NIHMSID: NIHMS527061; PubMed Central PMCID: PMC3871197.
- Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* 2005 Feb 24;33(4):e36. PubMed PMID: 15731329; PubMed Central PMCID: PMC549575.
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature.* 2002 Dec 5;420(6915):520-62. PubMed PMID: 12466850.
- Williams A, Spilianakis CG, Flavell RA. Interchromosomal association and gene regulation in trans. *Trends Genet.* 2010 Apr;26(4):188-97. PubMed PMID: 20236724; NIHMSID: HHMIMS194229; PubMed Central PMCID: PMC2865229.
- Williamson I, Eskeland R, Lettice LA, Hill AE, Boyle S, et al. Anterior-posterior differences in HoxD chromatin topology in limb development. *Development.* 2012 Sep;139(17):3157-67. PubMed PMID: 22872084; PubMed Central PMCID: PMC3413162.
- Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 2005 Jan;3(1):e7. PubMed PMID: 15630479; PubMed Central PMCID: PMC526512.
- Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, et al. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 2010 Nov 15;24(22):2543-55. PubMed PMID: 20966046; PubMed Central PMCID: PMC2975930.
- Yokota H, Singer MJ, van den Engh GJ, Trask BJ. Regional differences in the compaction of chromatin in human G0/G1 interphase nuclei. *Chromosome Res.* 1997 May;5(3):157-66. PubMed PMID: 9246408.
- Yusufzai TM, Felsenfeld G. The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A.* 2004 Jun 8;101(23):8620-4. PubMed PMID: 15169959; PubMed Central PMCID: PMC423244.
- Zhao H, Kim A, Song SH, Dean A. Enhancer blocking by chicken beta-globin 5'-HS4: role of enhancer strength and insulator nucleosome depletion. *J Biol Chem.* 2006 Oct 13;281(41):30573-80. PubMed PMID: 16877759.

Agradecimientos

Agradecimientos

Bueno pues ya casi está, todavía queda arreglar algunas cosas, muchas otras se quedarán tal cual, aunque siempre todo puede mejorarse. Pero el tiempo es finito y la vida está llena de cosas que hacer.

Y así es como terminan las buenas historias, dando las gracias. Siendo sincero, no ha sido una historia idílica, ha tenido momentos buenos, malos, muy buenos y muy malos, pero de todos ellos he aprendido algo. Lo importante, en realidad, son las personas con quienes he compartido esos momentos.

Gracias, Renata, por haberme introducido en este mundo, por enseñarme que la ciencia no tiene sentido si no la disfrutas, por tu coraje, valentía y actitud positiva ante las cosas y por ayudarme siempre que te he necesitado.

Gracias, Carmen y Yoli, por ayudarme a pasar mejor los malos momentos y disfrutar más de los buenos. Por vuestro cariño y vuestra ayuda y por los buenos ratos que me habéis hecho pasar. Gracias, Mario, por sacarme una sonrisa todos los días, ha sido un placer compartir estos años contigo. Por supuesto gracias, Anita, por compartir la jornadas interminables en el laboratorio, por haberme apoyado siempre, por haberme dejado compartir mis penas, por todas las risas que nos hemos echado, por tu gran amistad.

Gracias a los vecinos, María, Ines, Rocío, Javi, JR y Joaquin, por ayudarme siempre que lo he necesitado y por vuestro consejo.

Gracias a Rocio, Sol, Ana A, Xavi y todos los que habéis pasado por el animalario por hacerme la vida más fácil.

Gracias a los que ya estabais en el labo cuando llegué: Ana Neto, Royo, Miñan, Ozren, Silvia, Eli y Ze. Os tengo mucho cariño y me habéis enseñado muchísimo, cada letra de esta tesis tiene un poco de vosotros. Gracias, Juanito y Nacho, por haberos implicado en mi trabajo y haberme apoyado en los momentos más críticos, sé que mucho de lo que haya podido conseguir os lo debo a vosotros, sois dos grandes.

Gracias a los nuevos: Ensieh, Panos, Sandra, Marta y Sergio, por haber traído nuevas energías a labo, os deseo lo mejor, que trabajéis duro, que aprendáis mucho y que disfrutéis todo lo que podáis. Gracias, Rafa, por haber hecho mi última etapa mucho más llevadera, por habérmelo hecho pasar muy bien, por haberme ayudado muchísimo y por tu amistad. Eres una gran persona y estoy seguro de que te irá muy bien.

Gracias Antonio, por todos estos años de amistad. Gracias por todo lo que me has enseñado no sólo de la ciencia sino también de la vida, gracias de verdad por todo lo que me has ayudado. Para mi ha sido un placer compartir este viaje contigo.

Sobre todo gracias a ti, Jose Luis. No puedo sino quedarme con lo bueno porque ha sido mucho. Te agradezco la oportunidad que me has dado, de aprender y de formarme. Te agradezco el haber podido participar de unos proyectos tan extraordinarios de la mano de personas tan extraordinarias. Te agradezco que te hayas preocupado tanto por mi formación

y por llevar a buen puerto mi trabajo. Eres una buena persona y un gran profesional. No puedo sentirme sino orgulloso, de haber hecho mi tesis contigo.

Por último quiero dedicarle unas palabras a quienes más le debo el haber concluido este trabajo: a mi familia. A mis padres por su amor infinito e incondicional, por animarme siempre a seguir adelante, por enseñarme el valor del esfuerzo, del trabajo, del compañerismo y la solidaridad. No sabéis el profundo orgullo que siento por vosotros. Y por supuesto quiero darle las gracias a mi nueva familia. A Rocio, mi amor. Por tu paciencia más que infinita, por estar siempre a mi lado, por mostrarme una sonrisa cuando merecería un guantazo, por hacerme reír incluso cuando no quiero, por calmarme cuando estoy preocupado y por ponerme las pilas cuando me calmo demasiado, por hacerme vivir y por hacerme feliz. ¡Qué afortunado soy de haberme cruzado contigo!.

Y a ti Carlos, a ti con todo mi amor, te dedico este trabajo.

